S

Safety, Data

M. Murat Civaner
Department of Medical Ethics, Uludag University School of Medicine, Bursa, Turkey

Abstract

This entry discusses the importance of data safety, and the problems relating to (a) failure to maintain data safety and (b) abuse of data by state and commercial entities are discussed. The issues related to healthcare provision context are elaborated as well. Finally, some suggestions are made in order to improve the current practices and regulations on data safety.

Keywords

Data; Privacy; Confidentialty; Healthcare

Introduction

Gathering information and keeping records were always crucial for mankind in the endeavor to understand the people and the environment surrounding them. It has always been one of the most critical components of the struggle for survival. As our memory is limited and our vision is narrow at a given moment, we need data to comprehend the big picture and to make predictions and future plans accordingly. Along with the development of digitalization and Internet, now it is possible to gather data faster, wider, and larger more than ever. The possibility of coding all kinds of information as "1" and "0" enables us storing huge amount of data and transferring it swiftly.

Personal information might be in various forms and used for different purposes. It is not limited to name, age, or gender. It includes all information that identifies a certain person such as ethnic origin, faith, political affiliation, sexual orientation, social security number, fingerprints, e-mail address, IP numbers, hobbies, and relatives. It also covers personal health records, predictive analytics, data of health-related sensors, and gene sequencing technology (Liu et al. 2014). In the context of healthcare, recording personal information is a must to provide services adequately. Organizing a healthcare system, which allows healthcare workers to communicate more quickly and accurately, is only possible where the appropriate means are available. We use videotapes, audiotapes, photographs, DVDs, or hard disks to store personal and medical information, in addition to traditional instruments. Data are needed not just for diagnostic and therapeutic services but also implementing and monitoring public health programs for prevention and promotion, improving services and patient safety, biobanks, participating community in decision-making and audit mechanisms, and conducting researches to improve the services. This is where the new technologies come

H. ten Have (ed.), *Encyclopedia of Global Bioethics*,
DOI 10.1007/978-3-319-09483-0

to scene by providing new opportunities, not just for the local context but also on a global scale. Telemedicine, teleconsultation, teleradiology are increasingly used for consultation of diagnostic tests and treatment planning as they provide fast and convenient ways to communicate all around the world while decreasing the costs. As the Internet and mobile technologies develop, storing and transferring sensitive data increase parallel to the new concepts such as electronic health (e-health), mobile-health (m-health), and ubiquitous-health (u-health) as a recent development. The endeavor to digitize the information of human body including genetic materials, cells, organs, and even the brain transforms a human being herself into a kind of data.

However, various kinds of data provided by new technologies come with their own costs while creating advantages. Keeping data safe while recording, storing, transmitting, and disposing has been a vital issue as electronic media are becoming an indispensable part of daily life. Three main concepts mostly referred about securing the information are Confidentiality, Integrity, and Availability. "Confidentiality" is defined as "the assurance that information is not disclosed to individuals or systems that are not authorised to receive it," whereas "Integrity" is "the assurance that information can't be modified by those who are not authorised to modify it, or that any such modifications will not pass undetected" (UK Government 2015). As for the "Availability," it is described as "the assurance that information is available when it's needed, and that mishap or malice cannot affect the ability of systems to provide information when requested." These concepts point out the problematic areas of data safety and vulnerabilities as well.

Confidentiality, Privacy, Autonomy

The question of who controls the information does matter more than the question of who owns it (British Medical Association 2012). In that sense, breaches of confidentiality and privacy cause individual losing control on personal integrity and even her own body by transferring the power on self to the other entities. It could stigmatize her in the society for a lifetime, cause financial losses, and might be harmful by causing various disadvantages including discrimination. Maintaining the private information's control in the limits of personal will is paramount so that the person loses her freedom and autonomy otherwise. In the Internet era dissemination of information is now much faster, wider, and irreversible, therefore uncontrollable dissemination affects not just that particular person but also her family and acquaintances. Leakage of celebrities' private pictures from a cloud-based service is a good example of it.

Failure to maintain the data safety causes problems not just on the personal level but also on the societal level. One of the problematic areas where confidentiality, privacy, and autonomy are compromised is the citizen-state relationship. The organization created by society for the purpose of common good, which is called "state," is authorized to collect data for providing all kinds of public services starting by birth. It creates concerns about personal liberties as the state itself is the legitimized construct that holds the most intensified power in the society, and therefore its possible abuse by administrators and politicians might have gravely harmful consequences to society and individual. These concerns increase where the data about citizens are collected in the name of national security. Especially after the attacks to World Trade Center in USA in 2011, a new concept of security was created where the notion of "public security" was upgraded to a higher level and replaced by "national security." "Preventive war" became the central concept of national security argument against the enemies of state and regime. Using new technologies, states now has the ability to monitor and record their citizens' movements constantly, not just in the public places but in the virtual world as well. "Big brother" or Bentham's "Panopticon" are becoming a real phenomenon with surveillance networks, while more data provide more control. Ability to track people's movement was extremely increased so that a software developed recently is capable of predicting future behavior of a certain person by mining data from social

networking websites (The Guardian 2013). This invisible intrusion into private life of individuals causes a conflict between personhood rights and national security. Along with the oppression created by ideological state apparatuses as defined by Althusser (2006), an individual develops an autocontrol in herself eventually, and at the end, there would be no need to control her by exercising power externally. As Allmer put it clearly, "The Panopticon creates a consciousness of permanent visibility as a form of power, where no bars, chains, and heavy locks are necessary for domination any more" (Allmer 2012). Maintaining data safety seems crucial to avoid this "permanent visibility" status therefore protecting autonomy.

The other problem with failure to keep the data safe on societal level is that it provides novel possibilities of exploitations to commercial entities. Internet pages, e-mails, and cell phone messages provide faster and cheaper marketing bombardment and open up new doors to a nearly uncontrolled world of subliminal messages and hidden advertising. Lifestyle and personal preferences are monitored by spending patterns over time. Companies track down the customers with the help of mobile phone service providers, and advertisements or discount messages are immediately sent about a certain product which is sold in the store that customer stepped in. They trade and exchange the databases consisting of millions of customer's personal information, which were collected by phishing messages, product campaigns, or simply from each other. Data itself is becoming a valuable commodity not just for marketing purposes but also for the decisions of bank loans and insurance premiums. Facebook, for instance, in which people willingly write information about themselves, becomes one of the major data sources for companies in job hiring processes.

Data Safety in Healthcare

It is obvious that persons would not want their secrets disclosed without their consent, since unauthorized dissemination of personal information could be seriously harmful to that person in her family, in the society, or in her job. This risk is especially true when people get healthcare, preventive or curative, as they should give their most intimate secrets to healthcare workers correctly and accurately, if it is a necessity to provide the service they need. This is exactly the reason to make them more vulnerable in the relationship with medicine than the other parts of their life, in which they have more freedom to act and control. But in the context in healthcare, misery and pain caused by the health problem, knowledge asymmetry between them and representatives of medicine, education and language differences, all factors put them in a disadvantaged position limiting their liberty of giving personal information to others as they see fit. The only thing that could relieve them is the assumption that their information will be kept hidden, and that is why trust is fundamental for medicine.

Traditionally it is healthcare workers, especially physicians, who were defined as the responsible party to protect confidentiality and therefore to establish the relationship with the patient based on trust. However, in the new era of digitalization and healthcare reforms based on cost-effectivity, it is getting more difficult to honor this professional duty since the data are substantially beyond their control now. Ministry of Health, reimbursement institutions, insurance companies, drug stores, or companies in the medical industry would like to access data collected by physicians and healthcare institutions in order to decrease the costs, improve services, allocate resources, or invest in more profitable areas. Electronic media allows many people even in the hospital, including researchers, experts in quality management units, forensic medicine, administration, and data processing centers to access personal and medical information.

The collection of data in large databases and sharing them by vast networks increase the difficulty of protecting confidentiality, integrity, and accessibility. According to a recent study conducted in the registries of health data breaches in USA between 2010 and 2013, 6 breaches out of 949 reported breaches involved more than one million records (Liu et al. 2014). It was found that most occurred via electronic media, while

theft of electronic devices, hacking, and unauthorized access were the most frequent type of breaches.

New technological abilities create new ethical problems as usual. The question of who should be able to access which kind of data becomes an important issue. Physicians, for instance, might like to be authorized to access all information of all patients admitted to the hospital before for different health problems, including sexual orientation or HIV markers, whether or not their actual complaint is relevant. They might claim that they need all information to protect themselves and their patients, and if this information, especially about the communicable diseases, would not be available to them, they might claim that they would have a right to refuse to treat those patients. It creates a specific challenge in an atmosphere where professional values and patient rights are not protected as much as they deserve.

Another problem emerges within the trend of commercialization in healthcare. Companies that operate the data management systems of hospitals might claim that databases are their property therefore they have the right to have a copy and use it as they see fit after the contract is over. Disclosing the databases to commercial companies might be done by the health authorities of a country as well. By a recent policy called "care.data" in the UK (UK National Health Service 2015), it will be possible to use patient information for purposes other than their direct medical care and "the intention is to make it available – with some of the identifying information removed, but not always – to organisations outside of the NHS including universities, commercial companies, medical researchers and information intermediaries" (MedConfidential 2014). Furthermore, in another instance from Turkey, it was found out by the Turkish Court of Accounts that the Social Security Institution, the reimbursement body of the government, had sold the databases of all patients to five companies (Turkish Court of Accounts 2013).

It is possible to claim that basically two consequences emerge from the problems mentioned above. Informed consent, taken in order to respect right to self-determination, is losing its meaning in practice as the use of data for secondary and even tertiary purposes is possible, and the answer to the question of "who is the owner of data?" is getting blurred. Secondly, as the patients have less trust in medicine, they might feel that the only options they have are disclosing inaccurate or false information to physicians or not to visit them at all. It is obvious that it would be harmful both for their personal health and for the society, especially regarding communicable diseases.

Data as Intellectual and Commercial Property

The ownership of data is an important issue in production of knowledge as well. Regarding scientific researches, it is usually argued that maintaining data safety is vital in order to protect researchers' intellectual production and the value of their labor. In a knowledge production system where researchers usually compete with each other in isolated and uncooperative manner through academies and research institutes, data become a kind of property, albeit intellectual. Producing and keeping its ownership of that property bring income, possibilities for future grants, reputation, societal position, and "academic capital" as Bourdieu defined (Bourdieu 1984), while not having it or losing it brings otherwise. Therefore it is protected carefully until it is registered – licensed – and patented to the researcher by publishing it. Published data is protected as well, by copyright laws, patent laws, or intellectual property laws, where using it without permission or referring properly becomes a major crime in the context of intellectual production, called "plagiarism." Therefore, this is the reason why plagiarism and even theft are the main topics coming to mind when data safety in research context is discussed.

Various forms of plagiarism, copying datasets without permission while peer-reviewing, or violations of intellectual property rights are not rare and important problems indeed. However, there is another dimension of the problem which should be on the agenda of data safety discussions. The

concept of intellectual property rights is not used just for protection of single researcher's rights; rather, that concept is used to protect commercial interests. In current practice, life sciences researches are mainly sponsored by companies of medical industry; therefore companies claim that they have the right to keep the data without disclosing. Knowledge is transformed into a kind of commodity, where the data safety argument is used for hiding the data as companies put profits before society. Researchers have to sign contracts binding them not to release the data they produce in any form without the permission of the sponsor. According to a study aiming to identify the prevalence and determinants of data-withholding behaviors among 2,167 academic life scientists, one in every five participants had reported that "publication of their research results had been delayed by more than six months at least once in the last three years to allow for patent application, to protect their scientific lead, to slow the dissemination of undesired results, to allow time to negotiate a patent, or to resolve disputes over the ownership of intellectual property" (Blumenthal et al. 1997). It was found out by the same study that participation in an academic-industry research relationship and engagement in the commercialization of university research were significantly associated with delays in publication. Contracts forbidding disclosure create a pressure on researchers, and the cost of disclosing research results could be severe. Two examples of the most known cases are Dr. Nancy Olivieri and Dr. David Healy, who spoke publicly about potential dangers to patients found by their researches. They had personally experienced serious negative consequences, including getting fired from their positions and being exposed to damaging rumors to discredit them (Schafer 2004). Furthermore, in addition to contracts with individual researchers, companies secure their data ownership with international agreements such as The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) (World Trade Organization 1994). Thus legal right to hide the data becomes a usual part of data safety discourse, which legitimizes society losing control over it.

Measures to Keep Data Safe

Breaches and abuse of data cause mistrust to medicine, science, and state seriously, in addition to the harms given to individual and society. In order to protect the rights of vulnerables there are many regulations on national and international level including UNESCO Universal Declaration on Bioethics and Human Rights, World Medical Association Declaration on Ethical Considerations regarding Health Databases, European Council Convention on Human Rights and Biomedicine, and several legislations in the European Union (European Commission 2015a). Those regulations basically cover responsibilities to maintain confidentiality, integrity and accessibility, informed consent, and deidentification of data.

However it is not easy to claim that regulations provide sufficient protection, therefore data safety is continuing to be a critical problem. Implementing general principles of the regulations is getting more difficult as the speed and variety of technological advancements boost. Education of healthcare workers, medical students, contracted workers, and voluntary staff about the duty of confidentiality and practical rules of security is one of the points to improve. UK Department of Health suggests some security measures against theft inappropriate access by staff such as (British Medical Association 2012)

- Lock doors, offices, and filing cabinets.
- Avoid leaving paper or computer files open where they may be seen by others.
- Do not leave files unattended.
- Password-protect computer systems, and do not share passwords with other people.
- Change passwords at regular intervals to prevent anyone else using them.
- Always clear the screen of a previous patient's information before seeing another.
- Always log out of any computer system or application when finished.

It should be kept in mind that data safety should be handled holistically and that the

responsibilities of healthcare workers are just a part of it. Improving institutional and national policies are equally important, if not more. For instance, with the help of new possibilities it might be easier to reidentify anonymized data, therefore deidentification might lose its effectivity. Pseudonymisation might be useful as a second security layer. Encryption of data for all kinds of electronic storage and transmission might provide further protection. As for informed consent, its power of protection from misuse for secondary purposes is limited when the consent is blanket, since individuals lose its control; so it should be conditional instead of open ended. Patients should be informed about secondary usage, potential incorporation into aggregated databases, and who will be authorized to access data. Opting out from secondary usage should be an available option before and after giving consent. Public participation and accountability on how data are used should be a part of institutional and national policies, as it was suggested by Nuffield Council on Bioethics (Nuffield Council on Bioethics 2015): "Any data project should first take steps to find out how people expect their data to be used and engage with those expectations through a process of continued participation and review." People should be able to demand deletion for the records about themselves. The ruling of the Court of Justice of the European Union in 2014 on the "right to be forgotten" in relation to online search engines is a great improvement in that sense (European Commission 2015b). Finally, data as intellectual and commercial property should be a part of the debates on data safety, and intellectual property rights should be balanced against public rights and interests.

Conclusion

Data safety is an important problem in today's highly digitized world that increases security deficiencies inevitably. Breaches have a potential to affect individuals, institutions, even countries, and it could be claimed that actual regulations guided by the triad of "confidentiality, integrity, accessibility" which is classically recommended for information security are not sufficient enough to lessen concerns. Yet there are significant spaces for improvement regarding vulnerabilities on micro, mezzo, and macro levels.

In addition to the need of improving the measures aiming to protect data from unauthorized usage, there is a clear need to deal with the problems created by authorized usage in the era of Panopticon-state and commercialization. An individual's control over the data about her should be increased, and personal information should not be used for controlling or manipulating her by state and commercial entities. Public participation on every stage of data collection, storage, transmission, and disposal might be helpful to prevent abuses. Democratization would be a constructive step in order to guarantee that production and usage of data primarily aims at public interests, therefore re-establishes trust. This is also true for medicine, where trust relationship is crucial. Maintaining data safety and involving patients into decision-making about its usage seem essential for a better physician-patient relationship.

It is inevitable that we will live in more high-tech environments in future and electronic data will be a more inseparable part of us. It seems that control over data will be decisive on personal liberties and even our existence in society. From the perspective of human rights and professional values, advocating data safety in the name of improving human life and health will be the guiding point.

Cross-References

- ▶ Autonomy
- ▶ Biobanking
- ▶ Biolaw
- ▶ Biopiracy
- ▶ Biopolitics
- ▶ Biosecurity
- ▶ Censorship
- ▶ Citizenship
- ▶ Codes of Conduct
- ▶ Coercion
- ▶ Commercialism and HealthCare
- ▶ Commercialism in Scientific Research

- ▶ Committees: Research Ethics Committees
- ▶ Commodification
- ▶ Consent: Community
- ▶ Consent: Informed
- ▶ Cybernetics
- ▶ Digital Divide
- ▶ Discrimination, Ethics of
- ▶ Electronic Patient Records
- ▶ Epidemiology
- ▶ Exploitation
- ▶ Family Medicine
- ▶ Freedom of Speech
- ▶ Global Ethics
- ▶ Human Body
- ▶ Human Dignity
- ▶ Identity
- ▶ Information Technology: Ethics
- ▶ Integrity: Research
- ▶ International Documents
- ▶ International Law
- ▶ Internet
- ▶ Laboratory Science Ethics
- ▶ Managed Care
- ▶ Media Ethics
- ▶ Medicine and Ethics
- ▶ Military Ethics
- ▶ Nanotechnology
- ▶ Ownership
- ▶ Patenting
- ▶ Person
- ▶ Personalized Medicine
- ▶ Plagiarism
- ▶ Professional Ethics
- ▶ Professional-patient Relationship
- ▶ Research Policy
- ▶ Research: Animals
- ▶ Research: Clinical
- ▶ Research: Data Sharing
- ▶ Research: Embryo
- ▶ Research: Human Subjects
- ▶ Research: International Collaboration
- ▶ Research: Interspecies
- ▶ Respect for Autonomy
- ▶ Right to Health
- ▶ Risk
- ▶ Safety, Patient
- ▶ Scientific Misconduct
- ▶ Stigmatization
- ▶ Testing: Premarital
- ▶ Trust
- ▶ Vulnerability

References

Allmer, T. (2012). *Towards a critical theory of surveillance in informational capitalism*. Frankfurt am Main: Peter Lang.

Althusser, L. (2006). *On the reproduction of capitalism: Ideology and ideological state apparatuses (In Turkish)*. Istanbul: Ithaki Publishing.

Blumenthal, D., Campbell, E. G., Anderson, M. S., Causino, N., & Louis, K. S. (1997). Withholding research results in academic life science: Evidence from a national survey of faculty. *JAMA, 277*, 1224–1228.

Bourdieu, P. (1984). *Distinction. A social critique of the judgement of taste*. Cambridge: Harvard University Press.

British Medical Association. (2012). *Medical ethics today* (3rd ed.). West Sussex: Wiley.

European Commission. (2015a, July). Data protection legislation. Retrieved from http://ec.europa.eu/justice/data-protection/law/index_en.htm

European Commission. (2015b, July). Protection of personal data. Retrieved from http://ec.europa.eu/justice/data-protection/

Gallagher, R. (2013, February 10). Software that tracks people on social media created by defense firm. *The Guardian*. Retrieved from http://www.theguardian.com/world/2013/feb/10/software-tracks-social-media-defence

Liu, V., Musen, M. A., & Chou, T. (2014). Data breaches of protected health information in the United States. *JAMA, 313*, 1472–1474.

MedConfidential. (2014, November). Important changes to the way your personal data is handled. Retrieved from http://medconfidential.org/wp-content/uploads/2014/01/caredata_poster_green.pdf

Nuffield Council on Bioethics. (2015, February). Public participation should be at the heart of big data projects. Retrieved from http://nuffieldbioethics.org/news/2015/public-participation-heart-big-data-projects/

Schafer, A. (2004). Biomedical conflicts of interest: A defence of the sequestration thesis – Learning from the cases of Nancy Olivieri and David Healy. *Journal of Medical Ethics, 30*, 8–24.

Turkish Court of Accounts. (2013). Report on Social Security Institute (In Turkish). Retrieved from http://www.sayistay.gov.tr/rapor/kid/2013/Sosyal_Güvenlik_Kurumları/SOSYAL%20GÜVENLİK%20KURUMU.pdf

UK Government. (2015). Information security. Retrieved from https://www.gov.uk/service-manual/making-software/information-security.html

UK National Health Service. (2015, August). Your health and care records. Retrieved from http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Pages/care-data.aspx

World Trade Organization. (1994, April). The agreement on trade-related aspects of intellectual property rights. Retrieved from https://www.wto.org/english/tratop_e/trips_e/t_agm0_e.htm

Safety, Patient

Gerard Magill
Center for Healthcare Ethics, Duquesne University, Pittsburgh, PA, USA

Abstract

The debate in bioethics around patient safety deals with the pervasive problem of medical error. A traditional reaction to medical error in the past focused on blaming individual professionals. This punitive method was called the *professional sanctions model*. However, it was unsuccessful insofar as medical errors continued extensively. An alternative is called the *patient safety model*. This new approach focuses on the organizational systems and processes underlying medical error. The goal is to shift from being reactive and punitive to being proactive and positive in the sense of preventing error and enhancing safety. At the core of this new method is the root cause analysis of errors to identify the underlying systems and processes that create an environment in which mistakes occur. This results in fostering a safety culture that encourages organizations to identify mistakes, to prevent sentinel events, and to support patients and families who have been hurt. As a result, this approach calls for better national reporting mechanisms of medical error to enhance patient safety in healthcare across the world.

Keywords

Leadership; Medical error; Medical mistake; Patient safety; Professional sanctions; Organizational systems; Quality of care; Root cause analysis; Safety culture; Safety paradigm; Sentinel event

Introduction

The debate in bioethics around patient safety deals with the pervasive problem of medical error in healthcare. This is a global problem insofar as the extent of avoidable medical error is enormous, causing very large numbers of deaths and even higher numbers of serious injuries. For example, in the USA, a famous report was published by the Institute of Medicine in 2000 called *To Err is Human: Building a Safer Health System* (IOM 2000) – this will be referred to as the IOM Report. The report calculated that up to nearly 100,000 deaths were occurring as a result of medical error each year in the nation. These numbers were calculated based on previously published data – though subsequent research has suggested that the number could be much higher. The puzzling issue is why the general public has not become upset about the extent of the problem. The explanation is that medical errors and deaths occur to one patient and in one hospital at a time. The lack of a central reporting mechanism to track the problem has made it very difficult for the public to understand what is going on. If the public became aware of the extent of the problem, public debate and widespread advocacy would likely result. Following this IOM Report, many organizational systems were implemented in healthcare to ameliorate the problem. A similar approach has occurred worldwide insofar as healthcare is now more attentive to designing and implementing organizational systems to prevent medical errors and to promote patient safety. This analysis discusses this focus on patient safety in healthcare as a function of organizational ethics.

Background

Until the turn of the millennium, the general approach to medical error across the world was to blame the clinician for negligence or malpractice. Typically, hefty penalties accrued not only as a punishment to the individual concerned but also as a warning to other clinicians to avoid similar conduct. In one sense this accusatory and reactive

approach made sense insofar as a patient either died because of a medical error or suffered significantly – redress was sought through punishment to prevent recurrence. However, this seemingly plausible reaction had very little impact on reducing the extent of medical error in healthcare. Hence, another approach emerged that initially seemed odd insofar as it shifted the blame from the individuals causing the medical error. The blame shifted to organizational systems in healthcare that were construed as being faulty or inadequate, creating an environment that made medical error possible. The basic idea was to fix the systems underlying medical error rather than to fix blame on good clinicians who were caught up in problematic structures. Of course, this new approach does not ignore the continuing need for investigations into intentional or reckless behavior around which malpractice legitimately revolves.

This shift from fixing blaming on individuals to fixing underlying systems has occurred in healthcare across the world in an effort to protect patients from medical error. As a result, the focus on organizational systems that cause medical error has generated a new emphasis in organizational ethics upon patient safety in healthcare. To discuss this dramatic shift in healthcare globally, the analysis considers the varying types of medical error and the contrasting models to deal with the problem (Magill 2013, 2006; some of the concepts from these publications have been incorporated and developed for this contribution). The goal is to move away from a paradigm of heat and blame to a paradigm of light and remedy. This means avoiding a focus on blame where individual clinicians are accused of causing medical errors. The alternative is to shed light on the controversy by examining the root causes of medical error to implement organizational remedies that support patient safety.

Medical Error and the Safety Paradigm

It can be helpful to note that discussions about patient safety can transfer to safety of human subjects in medical research. There are well-known examples of patients dying in research protocols, such as Jesse Gelsinger who was the first patient to die in a research protocol on gene therapy. However, the purpose of this analysis is patient safety – considering the safety of patients, not research subjects or scientific misconduct. The ethical implications of medical error negatively impact not only patients by compromising their safety but also healthcare professionals and organizations by placing them in untenable situations. Medical error constitutes a foundational compromise of the basic tenets of professional standards of practice and of an organization's integrity insofar as patients are killed or injured. The problem for professionals and organizations is compounded when they try to cover up the medical error or explain it away to patients who may not be able to understand medical causality. As a result, the ethical debate around patient safety is not just about the quality of care but also about professional standards of practice and organizational integrity (Panesar et al. 2014). To engage the ethical debate around patient safety, it is necessary to have a sense of the extent of medical errors in healthcare.

Medical errors can occur in any stage of care from diagnosis to treatment, including preventive care. There are different types of errors, often clustered into medication errors, surgical errors, and diagnostic errors (Wachter 2012). There are many ways in which diagnostic errors occur: for example, there can be a mistake or delay in clinical diagnosis; there can be a failure to provide relevant or indicated tests; and there can be a failure to act on the results of tests or when using tests that are outmoded. Furthermore, there are many ways in which treatment errors occur: for example, during surgery; in avoidable delay in treatment; when there is inappropriate care; and when prescribing, dispensing, administering, and monitoring medications. Also, many other forms of medical error occur: for example, when there is communication failure or equipment malfunction, when there is fatigue among clinicians, when there is poor infection control, etc. The list of errors and their varying types goes on expansively.

This variety of medical error generates a broad definition of patient safety as freedom from

accidental injury. In general, the approach to medical error that is adopted here reflects the philosophy of error in the work of James Reason (1990, 1997). In this broad context medical error can be defined as the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim. Even though not every medical error actually harms a patient, a category of errors can be specifically designated to identify those that cause harm to patients. These are called preventable adverse events – an adverse event is an injury resulting from a medical intervention. For example, by some estimates nearly one-third of adverse drug events among outpatients can be prevented.

However, it is important to distinguish a medical error from a lapse or a slip that occurs when an action is not what was originally intended. That can be described as being an error of execution. However, it is crucial to note that a slip or lapse is not necessarily minor or unimportant because patients can die from them. In contrast, there is an error of planning when an action does not achieve the outcome that was intended because the intended action was mistaken.

An additional distinction needs to be drawn between latent errors and active errors. With regard to active errors, they typically occur with operators at the front line where they intersect with patients. As a result, the effects of active errors are immediate. With regard to latent errors, they typically are distant from an operator's control reflecting poor design in a system. There is a shift in discourse about medical error from focusing on the clinician who enacts an error (active error) to underlying systems that create an environment where an error can occur (latent error). This shift emphasizes that the greatest threat to a patient's safety must be associated with systems. It is understandable why a standard response to a medical error focuses on active errors as being enacted by a specific clinician. Nonetheless, it is now widely understood that the failure of systems is much more critical insofar as medical errors tend to be linked with latent failures. In other words, creating safe systems in the healthcare environment requires a focus on latent errors as system failures. At the heart of this approach is the recognition that a system is a set of interdependent elements interacting to achieve a common aim (Zipperer 2014).

This approach to error that focuses on reviewing and designing underlying systems to support safety has been extraordinarily successful in other industries. For example, there is a safety paradigm in the airline industry that can shed light on the importance of a systems approach to error and safety. Civilian airlines adopt a comprehensive approach to flight and passenger safety. They insist upon a variety of related interventions including the following: accident investigations and incident reporting, meticulous record-keeping, mandatory standards for safety, training programs, data safety and monitoring boards, and research for continuous quality improvement. As a result, a variety of regulatory bodies have been developed in different countries across the world to provide oversight for the safety of civilian flights. Also, a crucial component in the success of noticeably increased flight safety has been the establishment of a process for confidential incident reporting. By freeing civilian airlines from regulatory reprisals for reporting pilot error and near misses, pilot cooperation has soared and enhanced passenger safety has resulted. This safety paradigm can be applied to healthcare by shifting from focusing on individual culpability for error to emphasizing organizational systems for patient safety.

Individual Culpability Versus Organizational Systems

Efforts to reduce medical error in the past highlighted the individual culpability of a clinician for the mistake. The individual clinician was blamed for the medical error and punished in an effort to prevent future recurrence either by the guilty individual or by others. This was known as the *professional sanctions model*. The reactive model dwelt upon who was responsible for the error, accusing a professional of carelessness or incompetence, rather than trying to clarify what precisely caused the mistake. This model relied upon establishing a climate of fear and shame to

deter individuals from making mistakes. Not surprisingly, this approach generated a tendency to cover up errors and a failure to report them, thereby escalating the problem rather than diminishing it. The failure of this model is widely recognized insofar as it was unable to reduce the extent of medical errors. A different model was urgently needed to more effectively deal with avoiding medical error and promoting patient safety.

The model that has made much more effective progress to diminish error and enhance safety is known as the *patient safety model*. This requires an understanding of the connection between risk management, quality improvement, and patient safety (Dekker 2011). The patient safety model seeks changes in the systems and processes that create the environment in which practices occur that cause individuals to make medical mistakes. The purpose is to identify exactly what occurred in a medical error, getting to the root cause of the problem, in order to implement safeguards in the underlying systems and processes that will prevent recurrence. This proactive organizational approach to the problem of medical error encourages reporting of error to remedy the system glitches that cause it and thereby lead to significant quality improvement of patient care. At the core of this proactive and positive model is what is known as root cause analysis whereby the underlying causes of the mistake are identified to prevent recurrence.

Root Cause Analysis

The purpose of root cause analysis is to reduce medical error and foster patient safety by identifying the precise cause of error and by preventing system failures (Joint Commission 2003, 2012). This undertaking enables healthcare to adopt a proactive approach that focuses on the underlying systems and processes rather than a reactive approach that concentrates on the personal performance of an individual professional to assign blame for error. In other words, identifying the causal factors of an error is what matters. Part of the vocabulary associated with root cause analysis is that of sentinel events – here, the term "sentinel" indicates an unexpected occurrence or risk that requires immediate investigation. However, sentinel events are not the same as medical errors: not all sentinel events result from a medical error (Joint Commission 2008). By adopting a systems approach rather than focusing on individuals, root cause analysis focuses primarily on organizational processes rather than on personal performance. Hence, a successful root cause analysis does not assign blame, but enables a team to understand the causes or potential causes of medical error and the required changes that are likely to prevent their recurrence.

Generally, a root cause analysis of a sentinel event clarifies the systemic underpinning of a problem to diminish medical error and improve patient safety (Wachter 2012). This involves a variety of interrelated steps, including the following: identifying risk reduction strategies; designing, seeking consensus around, and implementing improvements to the systems and processes underlying the problem; developing outcome measures for continuous improvement in the quality of care; and reporting the results in a public manner.

More particularly, the process of root cause analysis has very specific steps that should be followed to maximize the reduction of medical error and the enhancement of patient safety. Before any sentinel event occurs, the healthcare institution should create an environment that is conducive to pursuing a root cause analysis in order to improve patient safety. This means that the healthcare institution should provide sufficient resources to ensure that a reliable investigation can occur in a timely manner when required. When a sentinel event is reported, a team should be assigned to assess the event as an undertaking that is designed to be objective and fair.

The focus should be on risk reduction by considering systems and processes and not upon blame assignment by accusing individual professionals. The problem should be studied in a comprehensive manner beginning with a precise account of the criteria being used and an accurate description of what occurred or what nearly happened – sentinel events include medical errors

and near misses, so to speak. The data should be gathered and recorded meticulously and in a manner that safeguards data protection as well as patient confidentiality. An example of the processes involved in undertaking a root cause analysis and a summary of sentinel event data in the USA for 2014 can be found on the website of the organization that accredits hospitals, The Joint Commission (http://www.jointcommission.org/sentinel_event.aspx).

Safety Culture

Insofar as root cause analysis reflects the basic shift from individual culpability to focusing on organizational systems, this dynamic new approach provides a robust and flexible instrument to foster a culture of patient safety in healthcare organizations (Wachter 2012). Several ethical guidelines can be identified to develop an organizational culture that seeks to reduce medical error and promote patient safety as healthcare priorities.

Leadership is required across the healthcare organization to develop a culture of safety. Here leadership includes clinicians, executives, and boards of directors as an organizational responsibility (Youngberg 2012). This leadership commitment requires a dynamic vision to advance the quality of patient care in which patient safety is a priority objective for the institution. This commitment also requires appropriate allocation of resources – these are needed to ensure effective team functioning for safety-related processes (recognizing that there tend to be fewer errors when team work is involved) and to establish clear oversight for patient safety within the organization. These processes include root cause analysis of sentinel events to identify actual and potential risks, to reduce the number and severity of adverse outcomes, and to support systems redesign as needed.

Effective leadership is manifest in fostering a learning environment across the organization that includes everyone in the endeavor to promote patient safety. There are several crucial components that are needed to develop a robust learning environment for patient safety. Training must be provided to encourage open communication to report sentinel events and to provide feedback from errors not only to explain why they occurred but also to implement improvement mechanisms and to track changes for effectiveness.

Effective leadership and a robust learning environment must acknowledge that human limits are inherent to any design process. Because of the inherent limitations that humans bring to any process that involves patient safety, it is crucial to avoid reliance on memory and on vigilance because each can be prone to serious compromise. To avoid reliance of memory, protocols and checklists should be designed and adopted. An example is the civilian airline industry where pilots of the aircraft work through checklists meticulously as they prepare for the flight. Reliance on vigilance should be avoided because of the limited nature of the human attention span – our brains are not computers. Hence, key processes should be simplified to minimize problem solving and reduce the likelihood of error; and work processes should be standardized enabling personnel to work safely. In addition to concerns with memory and vigilance, jobs should be designed with safety as the priority such as giving attention to work hours and assignment loads, staffing ratios, fatigue, or sleep deprivation. Also, being conscious of safety in the workplace should encourage the use of constraints and nudging functions to guide users to the next action or decision.

While safety design in the workplace is indispensable for a safety culture, the organization also needs to be continuously watchful in the sense of anticipating the unexpected. This attentiveness requires a proactive mindset that has two interrelated features. There needs to be an ongoing scrutiny of systems and processes to antecedently detect potential threats to safety. There also needs to be an ongoing design and redesign of systems and processes both to avoid accidents (such as by automating tasks that are repetitive, time consuming, and error prone) and to recover from accidents (such as by making errors visible when possible and making it easy to reverse operations).

Because of the pivotal significance of developing a culture of patient safety across healthcare organizations to diminish medical error, it is indispensable that accrediting processes for healthcare insist on this matter. In many countries this has led to the development of national patient safety goals that healthcare organizations are expected to achieve. Naturally, as more goals (e.g., for hospital care or for long-term care) are achieved each year, patient safety will be enhanced significantly (for an example, see http://www.jointcommission.org/standards_information/npsgs.aspx). These goals can be connected with the extensive variety of specific cases of medical error to develop basic competencies in patient safety (Johnson et al. 2015).

Conclusion

As progress on patient safety continues, organizational ethics seeks to encourage two new approaches: a compensation approach and a central reporting approach. If these gain widespread appeal, they will considerably enhance the agenda for patient safety. Talking with patients and families about medical error and injury is not an easy undertaking (Truog et al. 2011). However, the endeavor can become more productive in the context of an apology and compensation, especially if healthcare can give assurances that better reporting mechanisms are being developed.

On the one hand, there needs to be better systems to establish fair resolutions or settlements regarding medical errors. One widely acknowledged approach is for organizations to assume responsibility. This approach is called no-fault compensation. Here, compensation is provided to the victims of medical error without the organization accepting culpability or fault. This approach receives significant attention as an alternative to undertaking expensive lawsuits (Wachter 2012). These medical malpractice lawsuits occur in different ways in different jurisdictions around the world. In this compensation approach, healthcare providers adopt several measures after a medical error occurs: they explain to the patient (or the patient's representatives) the mistake's details and offer an apology for the harm done; they explain the interventions that have been put in place to prevent recurrence with other patients; and they offer compensation for the pain and suffering involved.

On the other hand, there needs to be better mechanisms for reporting medical error within institutions, between organizations, and across each nation. This will require a centralized system that mandates reporting of medical errors at a national level, preferably with confidential reporting systems included. If that data can be accrued in a reliable manner, the public will have a better sense of the problem and healthcare systems globally will be help more accountable for patient safety.

Cross-References

- ▶ Leadership, Ethics of
- ▶ Mistakes, Medical
- ▶ Public Debate
- ▶ Quality of Care
- ▶ Research: Human Subjects
- ▶ Scientific Misconduct

References

Dekker, S. (2011). *Patient safety: A human factors approach*. Bacon Raton: Taylor & Francis.

IOM. (2000). Institute of Medicine. *To err is human: Building a safer health system*. Washington, DC, National Academy Press. http://www.jointcommission.org/sentinel_event.aspx, http://www.jointcommission.org/standards_information/npsgs.aspx. Accessed 28 May 2015

Johnson, J., et al. (2015). *Case studies in patient safety: Foundations for core competencies*. Burlington: Jones & Bartlett.

Joint Commission. (2003). *Root cause analysis in health care* (2nd ed.). Oakbrook Terrace: Joint Commission Resources.

Joint Commission. (2008). *Understanding and preventing sentinel and adverse events in your health care organization*. Oakbrook Terrace: Joint Commission Resources.

Joint Commission. (2012). *Improving patient safety and worker safety*. Oakbrook Terrace: Joint Commission Resources.

Magill, G. (2006). Chapter 7. Ethical and policy issues related to medical error and patient safety. In S. McLean (Ed.), *First do no harm. Law, ethics, and healthcare* (pp. 101–116). London: Ashgate.

Magill, G. (2013). Chapter 24. Organizational ethics. In D. F. Kelly, G. Magill, & H. ten Have (Eds.), *Contemporary catholic health care ethics* (2nd ed., pp. 271–290). Washington, DC: Georgetown University Press.
Panesar, S., et al. (2014). *Patient safety and healthcare improvement at a glance*. Chichester: Wiley-Blackwell.
Reason, J. T. (1990). *Human error*. Cambridge: Cambridge University Press.
Reason, J. T. (1997). *Managing the risk of organizational accidents*. Aldershot: Ashgate.
Truog, R. D., et al. (2011). *Talking with patients and families about medical error: Guide for education and practice*. Baltimore: Johns Hopkins University Press.
Wachter, R. (2012). *Understanding patient safety* (2nd ed.). New York: McGraw Hill.
Youngberg, B. J. (2012). *Patient safety handbook* (2nd ed.). Burlington: Jones & Bartlett.
Zipperer, L. (2014). *Patient safety: Perspectives on evidence, information and knowledge*. Farnham: Ashgate.

Further Readings

Johnson, J., et al. (2015). *Case studies in patient safety: Foundations for core competencies*. Burlington: Jones & Bartlett.
Joint Commission. (2015a). *Patient safety systems and sentinel event policy*. See, http://www.jointcommission.org/sentinel_event.aspx. Accessed 13 May 2015.
Joint Commission. (2015b). *National patient safety goals*. See, http://www.jointcommission.org/standards_information/npsgs.aspx. Accessed 13 May 2015.
Panesar, S., et al. (2014). *Patient safety and healthcare improvement at a glance*. Chichester: Wiley-Blackwell.
Wachter, R. (2012). *Understanding patient safety* (2nd ed.). New York: McGraw Hill.
Zipperer, L. (2014). *Patient safety: Perspectives on evidence, information and knowledge*. Farnham: Ashgate.

Scientific Misconduct

David B. Resnik
Bioethics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA

Abstract

Scientific (or research) misconduct has become a global concern. This entry reviews famous cases of misconduct or alleged misconduct; definitions of misconduct; policies for reporting, investigating, and adjudicating misconduct; incidence and causes of misconduct; and strategies for preventing misconduct.

Keywords

Misconduct; Scientific research; Ethics; Integrity; Policy; Education

Introduction

In the last 30 years, scientific research has become increasingly global in scope. Following World War II, the USA, the former Soviet Union, the UK, France, and other European nations were the leading sponsors of scientific research. Today, the list of top research funders includes China, Japan, South Korea, Taiwan, India, Singapore, Australia, Turkey, Brazil, Iran, and other nations outside of the USA and Europe. Many research projects involve collaborations among scientists from different countries and scientific journals publish articles from all over the world.

Globalization affects all aspects of scientific research, including ethical conduct. In the 1980s and 1990s, ethical scandals in federally funded research in the USA made scientists and government officials aware of the need to address research misconduct. In the twenty-first century, scandals have arisen in many different countries outside the USA, including Canada, China, Denmark, Germany, Italy, Japan, the Netherlands, South Korea, and the UK. Scientific (or research) misconduct can compromise the integrity of the research record, erode the public's trust in science, and threaten public health and safety. It is important, therefore, for countries around the world to address research misconduct (European Science Foundation 2008; Ana et al. 2013; Resnik and Master 2013; Shamoo and Resnik 2014).

Some Well-Known Cases Involving Misconduct or Allegations of Misconduct

Research misconduct is by no means a contemporary problem. In 1830, British

mathematician, inventor, and philosopher Charles Babbage published a book titled *The Decline of Science in Modern England* in which he rebuked his peers for dishonest research conduct. Babbage distinguished between forging (making up data), trimming (removing data that is inconsistent with one's hypothesis), and cooking (designing an experiment such that it is not a genuine test of a hypothesis) (Shamoo and Resnik 2014).

One of the earliest misconduct cases involved museum curator Charles Dawson's discovery of skull bones in the gravel beds of Piltdown, East Sussex, UK, in 1912. Dawson claimed that the bones were fossils that provided evidence of a "missing link" between humans and apes. The discovery was controversial from the outset, and many scientists doubted that the fossil was genuine. The Piltdown man was proven to be a hoax in 1953 when recently developed chemical techniques showed that skull was a human cranium combined with the lower jawbone of an orangutan. The skull had been aged artificially to make it appear older than it was (Shamoo and Resnik 2014).

In 1974, William Summerlin was conducting skin transplantation experiments at the Sloan Kettering Institute in New York in Robert Goode's immunology laboratory. Summerlin's research involved transplanting patches of skin from black-haired mice onto white-haired mice. He claimed that culturing the tissue prior to transplantation lowered the risk of rejection. A technician who was cleaning the animals' cages discovered the black-colored patches of hair on Summerlin's white mice could be washed away with alcohol. The technician reported the finding to Goode, whom initiated an investigation. Summerlin admitted to fabricating data by using a felt-tip pen to draw patches of black hair on the white mice. The committee investigating the incident found that Summerlin had fabricated data in other experiments and required him to take a medical leave of absence. The scandal damaged Goode's reputation, even though he was not implicated in it (Shamoo and Resnik 2014).

In 1983, two science journalists, William Broad and Nicholas Wade, published a book, *Betrayers of Truth*, which raised awareness about fraud and deception in science. The book discussed the Piltdown and Summerlin cases and questioned the integrity of scientific icons, such as Galileo Galilee, Isaac Newton, and Gregor Mendel (Broad and Wade 1983). Broad and Wade argued that Robert Millikan, who won the Nobel Prize in Physics in 1923 for measuring the charge on an electron, had acted dishonestly. To measure the charge of an electron, Millikan dropped negatively charged oil droplets through positively charged plates. When a droplet was suspended in the air, the electrical force would be equal to the force of gravity. Millikan was able to determine the charge of an electron by calculating these forces. Historians who examined Millikan's laboratory notebook for these experiments found that he did not report 49 out of 189 observations (26 %) that were marked as "fair" or "poor" in his notebook, even though he said he reported all of his observations in the paper. Although Millikan's results have been validated many times by other scientists, Broad and Wade argued that his conduct was deceptive. However, others have argued that Millikan did not falsify data. He had a good understanding of his equipment and knew when it was working properly. He probably decided not to report observations resulting from experimental error. While he should have discussed issues pertaining to experimental error in his paper, he did not conduct fraudulent research (Shamoo and Resnik 2014).

One of the first well-known cases involving an international collaboration took place in the mid-1980s. The case involved Robert Gallo, from the National Cancer Institute (NCI) of the National Institutes of Health (NIH), and Luc Montagnier from the Pasteur Institute in France. The two investigators were working together on isolating a virus thought to cause acquired immunodeficiency syndrome (AIDS). Gallo and Montagnier exchanged cell lines they had been culturing in their laboratories, which they believed were infected with different strains of the virus, and they published papers on the human immunodeficiency virus (HIV) in the same issue of the journal *Science*. When genetic tests revealed that the strains from the different laboratories were nearly identical, Montagnier

accused Gallo of stealing his strain and passing it off as his own. An investigation of Gallo found that he did not commit misconduct. The most likely cause for the genetic similarity between the strains is that both cell lines had been infected by a third, vigorous strain in Montangier's laboratory. The US and French governments reached an agreement that named Gallo and Montagnier as codiscoverers of HIV and required the sharing of patent rights for HIV blood tests (Shamoo and Resnik 2014).

A case that had a significant influence on the development of US federal government policies took place in the 1980s at the Whitehead Institute, a research center operated by the Massachusetts Institute of Technology (MIT) and Tufts University. Nobel Prize winning molecular biology David Baltimore and five coauthors published a paper in the journal *Cell* in 1986 in which they claimed to show how to use gene transfer methods to induce immune responses in cells. The NIH funded the research. Thereza Imanishi-Kari was an assistant professor who had conducted many of the key experiments reported in the paper. Imanishi-Kari's postdoctoral student, Margot O'Toole, had trouble replicating the experiments, so she asked to see Imanishi-Kari's laboratory notebooks. When O'Toole could not reconcile the data recorded in the notebooks with the data reported in the paper, she accused Imanishi-Kari of fabricating and falsifying data. Internal investigations by MIT and Tufts found that there was no evidence of misconduct, but an investigation by the Office of Research Integrity (ORI), which oversees NIH-funded research, found that misconduct had been committed. A Congressional committee that was looking into fraud in NIH-funded research also investigated the case, which was reported on the front pages of the *New York Times*. In 1996, a federal appeals panel found that there was not sufficient evidence to prove the Imanishi-Kari had committed misconduct, and it overturned the ORI's finding. Imanishi-Kari admitted to keeping poor record keeping practices, but not to misconduct. Although Baltimore was never implicated in this case, the adverse publicity damaged his reputation. Testifying before a Congressional committee, Baltimore described the affair as a witch-hunt (Shamoo and Resnik 2014).

Another case involving an international collaboration came to light in 1993, when Roger Poisson, a professor of Surgery at the University of Montreal, admitted to fabricating and falsifying data for 99 patients enrolled in the NIH-funded National Surgical Adjuvant Breast and Bowel Project (NSABP), a large multicenter NCI study led by Bernard Fisher from the University of Pittsburgh. Poisson admitted to changing his patients' medical data so that they would qualify for the study and receive experimental treatment. The misconduct was discovered when some NSABP statisticians noticed some inconsistencies in Poisson's data. NSABP scientists reanalyzed that data after removing Poisson's data and found that his misconduct had no effect on the overall results (Shamoo and Resnik 2014). The University of Pittsburgh and the NCI accused Fisher of knowingly publishing fake data, but the ORI found that there was no evidence that Fisher had committed misconduct. Fisher's reputation was damaged as a result of these investigations and public disclosures, and he sued the NIH, the University of Pittsburgh, and ORI for defamation. The lawsuit was settled out of court in 1997 for $3 million. The case spurred efforts by the Canadian government to develop research ethics policies (Shamoo and Resnik 2014).

Another case involving international collaborations occurred in 2002, when an investigatory committee at Bell Laboratories found that Jan Hendrik Schön, a rising star in the fields of condensed matter physics and nanotechnology, had faked data in at least 17 publications. Schön had been publishing in top scientific journals, such as *Science*, *Nature*, and *Physical Review Letters*, at the unbelievable rate of one paper every 8 days. Dozens of his papers were retracted. Schön came to Bell Laboratories from the University of Konstanz in Germany. In 2004, the university withdrew Schön's Ph.D. after a committee found that data reported in his dissertation were also fraudulent (Shamoo and Resnik 2014).

In 2003, researchers accused Bjørn Lomborg, an adjunct professor at the Copenhagen Business School in Denmark, of scientific dishonesty

related to the publication of his book *The Skeptical Environmentalist* in 2001. Lomborg's book challenged the consensus view among scientists that human-caused climate change will have dire consequences for the environment, the economy, and society. The researchers argued that Lomborg had fabricated, misrepresented, and misinterpreted data in the book. The Danish Committee on Scientific Dishonesty ruled that Lomborg had committed scientific dishonesty, but the Ministry of Science, Technology, and Innovation overturned its ruling on the grounds that there was not sufficient evidence to support it, and the definition of dishonesty was too vague. In response to the Lomborg affair, Denmark developed new regulations that limit the scope of scientific dishonesty to fabrication, falsification, or plagiarism or other serious deviations from good research practice (Resnik and Master 2013).

A case that had reverberations across the globe took place in 2005. In 2004 and 2005, a research group led by Woo Suk Hwang, a professor at Seoul University in South Korea, published two papers in the journal *Science* reporting the derivation of human embryonic stem (HES) cell lines by therapeutic cloning. If confirmed, the finding would be a major breakthrough in stem cell science. In December 2005, the editors of *Science* received a tip from an anonymous informant that some of the images of the cell lines reported in the 2005 paper had been faked. Shortly thereafter, Sung Roh, a member of Hwang's team, told reporters that 9 out of 11 cell lines reported in the 2005 paper were fabricated. A committee at Seoul University began investigating Hwang's research and found that all of the data in both papers had been faked. Hwang's papers were withdrawn, and he resigned his position at Seoul University. In 2006, Hwang was convicted of fraud, embezzlement, and breach of bioethics laws, but his sentence was suspended. A committee from the University of Pittsburgh found that Gerald Schatten, a faculty member who had collaborated with Hwang, had no involvement in the data fabrication, but that he had neglected his responsibilities as an author by not reviewing carefully the data and manuscript (Shamoo and Resnik 2014).

Several factors made it difficult to investigate the Hwang case. First, Hwang was a national hero in South Korea and received tremendous support from the media, the government, and university officials. Many of Hwang's peers did not want to criticize his research because he was bringing a great deal of money and prestige to Seoul University. Second, South Korean universities did not have adequate policies or procedures for reporting or investigating misconduct. Scientists who suspected fraud feared retaliation if they made an allegation against Hwang. Suspicions were initially reported to the media, not to university officials. Third, deference to authority is part of the South Korean culture, and many subordinates did not want to challenge Hwang. Fourth, South Korean universities did not have adequate programs in place to teach students about research ethics. After the Hwang case, South Korea initiated a number of reforms to promote research integrity (Kim and Park 2013).

Another case that emerged in 2005 involved Eric Poehlman, a professor at the University of Montreal. Poehlman had previously held positions at the University of Vermont and the University of Maryland. An investigation by the University of Vermont and ORI found that Poehlman fabricated or falsified data on 15 federal grant applications worth $2.9 million and 17 publications. The justice department also brought charges against Poehlman for defrauding the federal government. Poehlman agreed to a comprehensive legal settlement that addressed criminal, civil, and administrative actions brought against him. Under the terms of the settlement, Poehlman agreed to serve 1 year and 1 day in federal prison, to be barred for life from receiving federal grants, and to pay $180,000 to the government for restitution. He also agreed to pay $16,000 to the lawyer of Walter Denino, the student whom accused him of misconduct after he became suspicious of changes that Poehlman had made to a data spreadsheet. Poehlman's papers were also retracted (Shamoo and Resnik 2014).

A case that had adverse impacts on the health of children unfolded over the span of more than a decade. In 1998, British surgeon Andre Wakefield published a paper in the journal *Lancet*

claiming that exposure to the measles, mumps, and rubella (MMR) vaccine caused 12 healthy children to develop autism and intestinal problems. Members of the anti-vaccine movement hailed the paper as definitive proof that vaccines cause autism and other health problems. As a result, vaccination rates in the UK and other countries declined significantly. In 2004, journalist Brian Deer published an article in the *Sunday Times* in which he accused Wakefield of failing to disclose that his vaccine research was supported by a law firm preparing a lawsuit against MMR manufacturers and of not obtaining ethics board approval for his study. In response to these allegations, the UK General Medical Council (GMC) investigated Wakefield and decided to revoke his license for acting dishonestly by not disclosing significant financial interests and for ordering risky medical procedures, colonoscopies, and lumbar punctures, without appropriate qualifications or ethics board approval. In 2010, *Lancet* retracted the autism paper. In 2011, Deer published an article in the *British Medical Journal* claiming that Wakefield had falsified data in the paper. Deer reviewed the medical records of the children in the study and found that normal pathology results were changed to colitis in nine cases, three children reported as autistic did not have autism, and five children who were reported as normal prior to exposure to the vaccine already had developmental problems. Wakefield, who has sued Deer and the *British Medical Journal* for libel, claims that he did not falsify any data. He continues to provide advice and support to anti-vaccine groups (Shamoo and Resnik 2014).

Defining Misconduct

The most important conceptual issue pertaining to research misconduct is how to define it. To help clarify this issue, it is useful to distinguish between misconduct as an ethical and legal concept. Research misconduct as an ethical concept is simply behavior that violates accepted ethical standards for research. Misconduct in this sense is wrongful or unethical behavior (Resnik 2003). While this definition may be useful for teaching students about research ethics and developing codes of conduct, it is nearly impossible to enforce because it is excessively broad and vague. Many organizations have decided to legally enforce some types of serious violations of research norms that they classify as misconduct. This type of misconduct is behavior that violates certain types of legal rules. These rules may include polices adopted by institutions, funding organizations, or journals or various regulations or laws. For example, someone who violates US federal misconduct rules may lose federal funding. Someone who violates a university's rules against misconduct may lose employment. Someone who commits misconduct may also be charged with fraud if their behavior meets the definition of this concept. Fraud is a legal concept that can be defined as causing harm by misrepresenting a matter of fact (Resnik 2003). Someone who commits fraud may be liable under criminal or civil law. Misconduct proceedings conducted by institutions usually fall under contract law, whereas proceedings conducted by federal agencies fall under administrative law.

It is also important to distinguish between misconduct and questionable research practice (QRP) and good research practice (GRP). Scientific behavior ranges from ethical conduct (i.e., GRP) on the one hand to highly unethical conduct (i.e., misconduct) on the other. QRPs fall somewhere in between these polar ends of the spectrum. QRPs are behaviors that are ethically suspect or controversial, but not widely recognized as highly unethical. Some examples of QRPs include: omitting data outliers from one's analysis without providing an adequate explanation, overstating the significance of one's results, merely acknowledging an individual in a manuscript even though they deserved authorship credit, failing to keep adequate research records, violating the confidentiality of peer review, and not disclosing a conflict of interest to a journal (Shamoo and Resnik 2014).

During the 1980s and 1990s, US federal agencies defined misconduct in research as fabrication of data, falsification of data, plagiarism (FFP), or other serious deviations from accepted scientific practices. After several years of debate, the US government dropped the "other serious

deviations" category on the grounds that it was too vague and difficult to enforce. Also, some serious deviations such as sexual harassment, theft, or violations of human or animal research regulations may be covered by other policies (Resnik 2003). The definition currently used by federal agencies defines misconduct as simply FFP. Misconduct must be committed knowingly, intentionally, or recklessly, and does not include honest error or scientific disagreement. Fabrication is making up data; falsification is changing or omitting data or misrepresenting research by manipulating materials or processes; and plagiarism is appropriating another's ideas, words, results, or processes without giving proper credit (Office of Science and Technology Policy 2000).

While the federal definition of misconduct has considerable influence, it is not universally accepted. Nearly 60 % of US universities have definitions of misconduct that go beyond FFP. Some of the other behaviors that universities classify as misconduct include serious deviations from accepted research practices, significant or material violations of regulations, misuse of confidential information, interfering with a misconduct investigation, inappropriate authorship, and misappropriation of property (Resnik et al. 2014). Other countries also have adopted definitions of misconduct that include behaviors other than FFP (European Science Foundation 2008). For example, the UK Research Council's (2012) definition of misconduct includes FFP and inappropriate authorship and failure to exhibit due care for human or research subjects; Canada's includes destruction of research records and mismanagement of conflicts of interests (Tri-Council Agency 2011); China's includes violating submitting false résumés (Zeng and Resnik 2010); and as noted earlier, Denmark's includes other serious deviations from good research practice.

Disagreements about how to define research misconduct could pose ethical and legal problems for international research collaborations, since a type of behavior treated as misconduct in one nation might not be in another. Collaborators from different countries might be unsure about how to deal with a type of behavior that is defined as misconduct in one place but not in another. One could try to deal with this issue by following local definitions (i.e., "when in Rome, do as the Romans do"), but situations might arise, such as unethical behavior in cyberspace or between nations, where locality would be unclear. To avoid confusions like this, the research community should seek to develop a universally recognized definition of misconduct. If this goal is not achievable, then scientists should try to develop a common core definition of misconduct (e.g., FFP) that would form the basis of other definitions. In the last few years, scientists and government leaders from around the globe have held conferences to discuss research integrity issues. A result of these efforts was the development of the Singapore Statement on Research Integrity in 2010. The Singapore Statement includes some useful ethical principles, but it does not define misconduct (Singapore Statement 2010).

Reporting, Investigating, and Adjudicating Misconduct

It is important to have fair and effective procedures for reporting, investigating, and adjudicating research misconduct to promote scientific integrity and protect the rights of the parties involved. The US government has developed policies for recipients of federal funding to follow. These policies, which have served as a model for others adopted throughout the world (Resnik and Master 2013), generally involve four stages: informal assessment, inquiry, investigation, and adjudication. During the first stage, someone suspects that misconduct has occurred and makes a report in writing to an institutional official (such as a department chair) who relays this report to the person in charge of research integrity, ethics, or compliance at the institution. The research integrity official will assess the report to determine whether the allegation fits the definition of misconduct and has some evidential support. If the research integrity official determines that the allegation fits the definition and has some evidential support, then he or she will appoint a committee to conduct an inquiry to determine

whether there is enough evidence to warrant an investigation. The inquiry committee may sequester and examine research records and interview witnesses. If the committee determines that there is enough evidence to warrant an investigation, then the research integrity official will appoint an investigation committee. This committee may also examine records, interview witnesses, and will send its findings to the research integrity official. When the official receives the findings of the investigation committee, he or she will decide how to adjudicate the matter. If the committee finds there was no misconduct, then the matter is over. If the committee makes a finding of misconduct, then the research integrity official must decide what sort of punishment to administer (e.g., termination of employment, supervision of research, education/training). If the research has been funded by a federal agency, the official will send a report to the agency to review. The agency may accept the report, require additional evidence or deliberation, or even conduct its own investigation. If the agency finds that there was no misconduct, then the matter is ended. If the agency finds that misconduct has occurred, then it may impose sanctions, such as denial of federal funding for a period of time. It will also publish an official finding of misconduct that will be made available to the public (Shamoo and Resnik 2014).

There are several important points to note about misconduct proceedings. First, misconduct proceedings are supposed to be kept confidential to protect the rights and reputations of the accused and other parties. As seen in some of the cases discussed above, confidentiality is, unfortunately, not always maintained, and the reputations of innocent parties have been damaged. Second, defendants have rights to due process that must be respected. They have a right to seek legal counsel, to question witnesses, and to appeal decisions. Sometimes witnesses and those who report misconduct (i.e., whistleblowers) also hire attorneys. Third, whistleblowers should be protected against retaliation. Unfortunately, whistleblowers sometimes suffer adverse consequences from their actions. Those who do not experience direct retaliation may develop a reputation as a troublemaker and have difficulty finding work or lose funding if their supervisor is found to have committed misconduct. To continue their scientific careers, whistleblowers may need to find new supervisors or transfer to different institutions. Fourth, to provide additional protection for whistleblowers, some institutions permit anonymous misconduct allegations. However, it may not be possible to maintain anonymity if the allegation leads to an inquiry or investigation, since the accuser may need to provide testimony. Fifth, misconduct proceedings can consume a great deal of time, money, and energy. As mentioned above, the Imanishi-Kari case lasted 10 years. Sixth, if researchers have published a paper that has been impacted by misconduct, they should print a retraction if the results are no longer valid or a correction if the misconduct resulted in a minor error. Publishing a retraction or correction helps to protect the integrity of the research record (Shamoo and Resnik 2014).

Incidence and Causes of Misconduct

A number of different studies have attempted to estimate the incidence of research misconduct. Estimates from surveys that ask respondents to report if they have direct knowledge of misconduct vary from 3 % to 32 % (Shamoo and Resnik 2014). A problem with these types of studies is that they may overestimate the misconduct rate because respondents may not have good evidence that misconduct has occurred and two different respondents may report the same incident on a survey. Another way of estimating the rate of misconduct is to ask researchers to self-report. A survey of over 3,000 NIH-funded scientists published in the journal *Nature* found that 0.3 % admitted to falsifying or cooking research data in the last 3 years (Martinson et al. 2005). A systematic review and meta-analysis of survey research found that 1.7 % of scientists had admitted to fabricating or falsifying data at least once (Fanelli 2009). A problem with self-reporting surveys is that they may underestimate the

misconduct rate because respondents will not want to admit to engaging in unethical or illegal activity, even if their confidentiality is protected (Shamoo and Resnik 2014). While the incidence of misconduct is probably quite low, it is still a significant concern for researchers, because it has wide-ranging adverse impacts on science and society.

The causes of misconduct can be divided into individual and environmental factors. Individual factors include the desire for success, money, prestige, or career advancement; financial interests; psychological stress and illness; and lack of moral character. Environmental factors include externally imposed pressures to produce results, poor supervision, and oversight of research; cultural variations pertaining to the conduct of research; institutional corruption; poorly managed conflicts of interest; and inadequate policy development and research ethics education programs. Although some researchers believe that only scientists who are mentally unstable or amoral would commit misconduct, evidence suggests that misconduct often occurs when good scientists succumb to pressures to cut corners or bend or break the rules (Shamoo and Resnik 2014).

Preventing Misconduct

The most important strategy for preventing misconduct is to educate students, trainees, staff, and faculty in the responsible conduct of research (RCR). Since the late 1989, the NIH has required graduate, postdoctoral students, and trainees supported by NIH funds to receive instruction in RCR. In 2009, the National Science Foundation (NSF) also began requiring students supported by NSF funds to receive RCR instruction (Shamoo and Resnik 2014). Other countries have also begun to implement RCR educational programs (Ana et al. 2013; Resnik and Master 2013). Education should address not only avoiding misconduct but also a variety of others topics related to good scientific practice, such as data management, record keeping, collaboration, authorship, mentoring, laboratory safety, publication, peer review, conflict of interest, research with human and animal subjects, and social responsibility (Shamoo and Resnik 2014). Educational activities should provide information about concepts, principles, and policies pertaining to research ethics and include discussions of cases. Instructional programs may consist of semester-long classes, workshops, conferences, lectures, and online learning modules. Individual mentoring can also play a key role in RCR education (Shamoo and Resnik 2014).

Another important strategy for preventing misconduct is to develop institutional policies that inform students, staff, and faculty about the expected standards of behavior. Institutional policies should deal with misconduct as well as the other research ethics topics mentioned above and should be publicly accessible. While many universities around the world have already taken steps in this direction, many have not, so further policy development is necessary. Research sponsors and scientific journals should also continue to develop policies that address misconduct and other ethical concerns.

Ethical leadership also plays a key role in preventing misconduct. Institutional leaders include laboratory directors, department heads, deans, vice presidents, and other people involved in the management, supervision, and oversight of research. Leaders can set a positive tone for the organization by modeling ethical behavior and expressing a commitment to ethical values and principles. Leaders who set a negative tone may encourage moral indifference and corruption. Some of the worst scandals in science and business have been the result of unethical leadership (Shamoo and Resnik 2014).

Finally, auditing of research records can help to prevent misconduct. Auditing can be helpful in detecting errors and inconsistencies in research, as well as deliberate violations of laws and institutional policies. Auditing is a standard practice in banking, health care, insurance, air travel, and many other industries. Audits can detect problems that people are not aware of or are not willing to report. Audits can be conducted randomly or for cause (i.e., when a problem emerges) (Shamoo and Resnik 2014).

Conclusion

Research misconduct is a global problem that threatens the integrity and trustworthiness of science and can have negative impacts on society. Scientists, government officials, research sponsors, and journal editors should take steps to prevent misconduct and minimize its impact on science. Some of these steps include adopting a universally recognized definition of misconduct; formulating policies and procedures for reporting, investigating, and adjudicating misconduct; and implementing educational programs in research ethics.

Acknowledgments This entry is the work product of an employee or group of employees of the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH). However, the statements, opinions, or conclusions contained therein do not necessarily represent the statements, opinions, or conclusions of NIEHS, NIH, or the US government.

Cross-References

- ▶ Animal Ethics
- ▶ Animal Research
- ▶ Animal Rights
- ▶ Biolaw
- ▶ Children: Research
- ▶ Codes of Conduct
- ▶ Commercialism in Scientific Research
- ▶ Committees: International Ethics Committees
- ▶ Committees: Research Ethics Committees
- ▶ Conflict of Interest
- ▶ Cultural Diversity
- ▶ Dual Use
- ▶ Education: Goals
- ▶ Engineering Ethics
- ▶ Ethics
- ▶ Global Ethics
- ▶ Human Rights
- ▶ Integrity: Professional
- ▶ Integrity: Research
- ▶ International Documents
- ▶ International Law
- ▶ Law and Morality
- ▶ Moral Relativism vs Universalism
- ▶ Organizational Ethics
- ▶ Professional Ethics
- ▶ Publication Ethics
- ▶ Research: Animals
- ▶ Research: Clinical
- ▶ Research: Data Sharing
- ▶ Research: Human Subjects
- ▶ Research: International Collaboration
- ▶ Whistle-Blowing

References

Ana, J., Koehlmoos, T., Smith, R., & Yan, L. L. (2013). Research misconduct in low- and middle-income countries. *PLoS Medicine, 10*(3), e1001315.

Broad, W., & Wade, N. (1983). *Betrayers of the truth: Fraud and deceit in the halls of science*. New York: Simon and Schuster.

European Science Foundation. (2008). *Stewards of integrity: Institutional approaches to promote and safeguard good research practice in Europe*. Strasbourg: European Science Foundation.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE, 4*(5), e5738.

Kim, J., & Park, K. (2013). Ethical modernization: Research misconduct and research ethics reforms in Korea following the Hwang affair. *Science and Engineering Ethics, 19*(2), 355–380.

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature, 435*(7043), 737–738.

Office of Science and Technology Policy. (2000). Federal research misconduct policy. *Federal Register, 65*(235), 76260–76264.

Resnik, D. B. (2003). From Baltimore to bell labs: Reflections on two decades of debate about scientific misconduct. *Accountability in Research, 10*(2), 123–135.

Resnik, D. B., & Master, Z. (2013). Policies and initiatives aimed at addressing research misconduct in high-income countries. *PLoS Medicine, 10*(3), e1001406.

Resnik, D. B., Neal, T., Raymond, A., and Kissling, G. (2014). Misconduct definitions adopted by U.S. research institutions. *Accountability in Research* (in press).

Shamoo, A. E., & Resnik, D. B. (2014). *Responsible conduct of research* (3rd ed.). New York: Oxford University Press.

Singapore Statement on Research Integrity. (2010). Available at http://www.singaporestatement.org/statement.html. Accessed 5 Mar 2014.

Tri-Council Agency. (2011). Tri-Council Agency framework: Responsible conduct of research. Available at http://www.rcr.ethics.gc.ca/eng/policy-politique/framework-cadre/#311. Accessed 5 Mar 2104.

U.K. Research Council. (2012). The research ethics guidebook. Available at http://www.ethicsguidebook.ac.uk/Research-Council-funding-122. Accessed 5 Mar 2014.
Zeng, W., & Resnik, D. (2010). Research integrity in China: Problems and prospects. *Developing World Bioethics, 10*(3), 164–171.

Further Readings

Redman, B. K. (2013). *Research misconduct policy in biomedicine*. Cambridge, MA: Massachusetts Institute of Technology Press.
Wells, F., & Farthing, M. (Eds.). (2008). *Fraud and misconduct in biomedical research* (4th ed.). London: Royal Society of Medicine Press.

Sexual Ethics

Anton van Niekerk
Centre for Applied Ethics, Stellenbosch University, Matieland, South Africa

Abstract

There can be no doubt that sexuality is one of the most powerful forces operative in human experience. While sex facilitates the process of reproduction, in human existence sex has developed into considerably more than this. Sex also represents one of the fundamental drives that constitute human behavior. In as far as ethics is the outcome of reflection on the rightness or wrongness and goodness or badness of all human behavior, sex has since the earliest origins of our culture been a theme of consistent moral deliberation, regulation, and even legislation. This entry firstly outlines aspects of the history of reflection on sexual matters. The approach in terms of natural law (also inspired by religious concerns) is discussed, leading up to the approaches of St. Thomas Aquinas in the thirteenth century and Immanuel Kant in the eighteenth century, as well as the Victorian guilt morality that dominated the nineteenth and the earlier twentieth centuries. Following this, the liberalized understanding of and reflection on sex, mainly precipitated in the second half of the twentieth century in the aftermath of the Second World War, receive attention. In the third section of the entry, a brief conceptual clarification of the term "sexual activity" is offered in light of the conceptual difficulties that a proper understanding of this term has yielded. The final part of the entry explores the way in which the understanding and practice of, as well as the reflection on, human sexuality have changed. The differences between St. Thomas Aquinas's and Thomas Nagel's notions of "sexual perversion" serve as the point of departure. The entry concludes with an emphasis on the role of sex in the understanding and practice of intimate human relationships.

Keywords

Sexuality: Sexual relations; Procreation: Homosexuality; Intimacy: Masturbation; Sexual morality: Sexual perversion

Introduction

There can be no doubt that sexuality is one of the most powerful forces operative in human experience generally. The reason for this has, in all probability, fundamentally to do with the function of sexuality in human procreation. People (and most, though not all, animals) procreate on the basis of sexual intercourse – a process that directly facilitates the fertilization of female zygotes (eggs) by male zygotes (sperm), resulting in embryos from which all humans develop to maturity. Nowadays we know (as will also be argued later in this contribution) that sex in human interaction (particularly in human intimate relationships) plays a more extended and complicated role than that of merely facilitating reproduction. No single reason fully explains the force and prominence of the sexual impulse in human relationships. What cannot be doubted, however, is that in as far as ethics is the outcome of reflection on the rightness or wrongness and goodness or badness of all human behavior, sex has since the earliest origins of our culture been a consistent theme of moral deliberation, regulation, and even legislation – for a long time indeed the most important or prominent such theme.

There was a time when ordinary, nonintellectual people's understanding of and association with the term "ethics" was fundamentally informed by one or more aspects of the phenomenon of sex. The immense influence of Judaism and Christianity in Western societies (and Islam in Mideastern culture) had the effect that sex out of wedlock, i.e., sexual activity that occurs outside the confines of a societally and publicly recognized marriage relationship between one man and one woman (sometimes, indeed, more than one woman), generally became acknowledged as the prototype of all "sin" (i.e., behavior contrary to God's will for humanity).

This idea was broadly reinforced by the ancient Greek philosophers, particularly Plato, whose ideas had a profound influence on some of the most influential early Christian thinkers, particularly St. Augustine. Plato, in the well-known speech of Pausanias in *The Symposium*, argues that sex is "in itself neither good nor bad" and can indeed be a good thing in the context of "heavenly love" (as opposed to "common love"). In this speech, Pausanias rejects the morality of the widespread custom in Athens for mature men to have sexual relationships with young boys. His plea is that men should rather wait until they are able to "form a lasting attachment and a partnership for life" with an adult (Plato 1951, 182a). However, Plato's general disparagement of the body and the desires of the flesh vis-á-vis the powers of the intellect (Plato 1961, 44d–47d) reinforced Christianity's traditional rejection of the moral value of sex (Freeman 2002, pp. 277–312).

Irrespective of the moral status (or lack thereof) that sexual relations have always had in human culture, the phenomenon of sex and its impact on human behavior, particularly during and in the aftermath of puberty, acknowledge and demonstrate the extent to which our lives and self-understanding as human beings are conditioned by our bodiliness or corporeality. We are what we are on the basis of our bodies. There is no coherent idea or representation of human identity independent from the reality of the human body.

The French phenomenologist Maurice Merleau-Ponty argues persuasively in his well-known book *Phenomenology of Perception* (Merleau-Ponty 1962) that our bodies, even though they are chemically constituted by the same elements that occur elsewhere in the universe and are subject to the same laws of nature, are never merely "things among other things." Our bodies are "lived-through"; they are centers with reference to which we perceive and experience whatever the world has to offer. As such, they are centers of the constitution of meaning. All meaningful experience is related to the perception of the body. The world unfolds as a meaningful, orientated space around the body as meaning-constituting point of reference. Whatever we perceive, we perceive as "bigger" or "smaller," "nearer" or "further," and "higher" or "lower," and all of these dimensions – so fundamental to making sense of the perceived world – only attain meaning with reference to the situatedness and spatial orientation of the human body.

Merleau-Ponty strikingly points out how we attain a pre-theoretical familiarity with the world – a familiarity that we are not (yet) able to understand and describe in theoretical terms – purely on the basis of being bodies in the world. We even attain some forms of knowledge (like the ability to balance ourselves on a narrow ledge, or to drive a car, or to play a musical instrument) in such a way that what we do when engaging in these activities is no more than the expression of pure bodily prowess (the "thinking," "knowing," and "doing" body) and not the outcome of intellectual, theoretical knowledge acquisition, deliberation, or formulation.

One of the aspects of this pre-theoretical bodily experience (or "practico-gnosis," as Merleau-Ponty calls it) is to become aware of our sexuality – normally (and gradually) at the age of puberty, although the phenomenon of child sexuality also occurs. Then – often as an unforeseen surprise – the world starts attaining a "sexual ambience"; we start recognizing people primarily in terms of their sex; we become much more aware of our own and other people's bodies, we start experiencing sexual desire, we seek and cherish friendships on the basis of sexual attraction, and the like. The remarkable aspect of this phenomenon is, again, that it is not the outcome of

conscious decisions; it is simply a new experiential orientation brought about by our bodily existence, without us deciding for it to happen or making any conscious choices that precipitate the experience. The world starts attaining a "sexual meaning" simply because, and on the basis of, the fact that our bodies constitute the mode by means of which we are what we are (Merleau-Ponty 1962).

In the rest of this entry, attention will firstly be paid to aspects of the history of reflection on sexual matters. The approach in terms of natural law (also inspired by religious concerns) will be discussed, leading up to the approaches of St. Thomas Aquinas in the thirteenth century and Immanuel Kant in the eighteenth century, as well as the Victorian guilt morality that dominated the nineteenth and the earlier twentieth centuries. Following this, the liberalized understanding of and reflection on sex, mainly precipitated in the second half of the twentieth century in the aftermath of the Second World War, will receive attention.

In the third section of the entry, a brief conceptual clarification of the term "sexual activity" will be offered in light of the conceptual difficulties that a proper understanding of this term has yielded.

In the final part of the entry, the way in which the understanding and practice of as well as the reflection on human sexuality have changed will be discussed. The differences between St. Thomas Aquinas's and Thomas Nagel's notions of "sexual perversion" will here serve as the point of departure. The entry will be completed with a conclusion that will stress the role of sex in the understanding and practice of intimate human relationships.

The History and Development of Reflection on Sexual Ethics

Alan Soble (n.d.) remarks that the history of reflection on sexual ethics has, broadly speaking, been characterized by positions of "metaphysical pessimism" which, only relatively recently, have been succeeded by positions of "metaphysical optimism." This distinction tries to capture the general measure of appreciation and moral commendation that sexual activities attained in the work of the most influential theorists in this field over the past two millennia.

The metaphysical pessimists about sex are the thinkers who, in one way or another, are in agreement that the phenomenon of sexual interaction somehow does not befit the dignity of the human person. These thinkers do, of course, realize that sexual intercourse is (or, at least until about 30 years ago, used to be) a necessary condition for human procreation. Sex ought, however, only to be morally tolerated because of this eventuality. The only context within which sexual activity (where the latter is only to be understood as heterosexual vaginal intercourse) is morally acceptable is within the institution of marriage. Marriage, in addition, is understood as the publicly recognized, ecclesiastically sanctioned, and unbreakable bond between one man and one woman. This is believed in spite of the Christian tradition's reverence for the Bible – a document that in its first part (the Old Testament) harbors (apparently without any moral indignation) a number of stories in which the central characters (Abraham, David, and Solomon) are involved in polygamous marriages.

Christianity, as we know, was the dominating influence in the societal structuration and ideological orientation of the Middle Ages and has maintained much of this influence until well into the advent of the modern world since the seventeenth century. Until recently, Christianity was also the dominant exponent of the pessimism about sex that Soble (n.d.) writes about. No thinker was more influential in this regard than St. Augustine of Hippo (354–430 AD). His ideas emanate from a deep pessimism about the human condition in its natural state. That condition is fundamentally characterized by "original sin" – the destructive consequence of the catastrophic act of disobedience to God of the first humans in Paradise and the consequent metaphysical curse on all their descendants (i.e., humanity in its entirety). This doctrine was originally construed by St. Augustine and became commonplace in the Christian tradition of the Middle Ages and thereafter.

St. Augustine regards sexual awareness as one of the first consequences of sin, hence Adam and Eve's immediate covering of their genitals after the fall from grace. St. Augustine also suggests that the curse of original sin is carried forward through all the generations by the very act of sexual intercourse (Augustine 2009). Without liberation from an external source, humankind is eternally damned. Our only hope of liberation and eventual salvation is through divine intervention as it occurs in the redemptive incarnation of God in the historical figure of Jesus Christ (God become human), manifested in his death on the cross and his resurrection from the grave.

Apart from this metaphysical theory which has lost most of its credibility in our time, there have been a number of other reasons for the pessimism about sex. It is often argued that sex objectifies people; the sexual act essentially consists of using another person for the sake of attaining sexual gratification. This is particularly taboo, not only in the Christian tradition but also for important thinkers of the Enlightenment such as Immanuel Kant. One of the best known formulations of his "categorical imperative" is "Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end" (Kant 1948, p. 96). The objection coming from Kantian ranks is that sex causes people to take an interest in other people mainly for the sake of the pleasure that they might yield and not for the sake of relating to the other person as a person.

As is now apparent with reference to Kant, the pessimistic view of sex is not confined to thinkers who associate with the religious sentiments of traditional Christianity, Judaism, and Islam (this entry does not have space to discuss the sentiments of the latter two traditions – sentiments that are by and large similar, if not more conservative and pessimistic, than those of Christianity). Another argument against the moral acceptability of sex has to do with the way in which sexual activity seemingly blurs the alleged boundaries between human and animal identity. Roger Scruton argues that, when partaking in sex, humans seem to lose control over themselves. That constitutes a threat to the integrity of personhood. In the sexual act a person yields to the desires, whims, and preferences of another and thus becomes subservient to what another person wants, even sometimes at the cost of yielding one's basic human dignity. Says Scruton: "In desire you are compromised in the eyes of the object of desire, since you have displayed that you have designs which are vulnerable to his intentions" (Scruton 1986, p. 82).

The pessimistic view therefore suggests that sex is a threat to the maintenance of human reason in behavior. It therefore only has a place in marriage, and there it ought to mainly function for the sake of procreation. As a general trend this attitude toward sex prevailed in intellectual circles up to the middle of the twentieth century, although exceptions certainly occurred. The nineteenth century, in particular, is remembered and often referred to as the high noon of conservative sexual bigotry and is often referred to as the epoch that embodied "Victorian" morality. The name is derived from the famous British queen whose moral disposition in this regard consisted of an unusual mix of deep conservatism and marked naivety. The story is told that, when homosexuality was statutorily outlawed in her time and she was requested to sign the bill, she initially refused because of the fact that lesbianism was included in the proposed law, and she obstinately refused to believe that a practice such as lesbian sex did or could exist! Victorian morality is characterized by a pronounced public rejection of promiscuous, out-of-wedlock sex in public, but a widespread participation (mainly by men) in covert sexual indulgences.

The aftermath of the Second World War since 1945 represents the time when the metaphysical pessimism of the tradition was for the first time comprehensively challenged and in many respects significantly transcended. The rise of the women's liberation movement which insisted on women's autonomy over their bodies and their right to enjoy sex as much as men played a role in this development as did the rapidly growing acknowledgment and institutionalization of human and civil rights and the almost universal rejection of censorship in liberal and social democracies. Pornography became publicly available and commercially accessible for the first time, as did displays

of not only nudity but of simulated and actual sexual activities in the sphere of films, theater, and other forms of public entertainment. An entirely new sexual morality came about which could be called a "sport and pleasure morality" in contradistinction to the "guilt and secretive morality" of earlier times.

In this newer, "optimistic" morality, the positive value of sexual activity in human relations as well as its functionality in providing pleasure for its own sake is increasingly acknowledged. Sex is seen as a bonding mechanism, something that contributes directly to the quality of intimate relationships. This is a cultural environment where the idea that pleasure is intrinsically good is recognized and propagated. Pleasure – particularly the kind of pleasure that sex causes – does not only have instrumental value, i.e., the value of facilitating human procreation. Liberated from the metaphysics of both the religious and the natural law interpretations of the nature and purpose of sexuality, the necessary link between sex and marriage is increasingly questioned and denied. Irving Singer in this respect points out that sex is indeed a drive, but then a drive distinctly different from the other drives which we share with the animals. It is much rather an "interpersonal sensitivity"; it is not that which degrades us to the level of animal behavior, but exactly that which elevates our experience of intimacy to a level that no animal could ever experience. Singer writes: "Though at times people may be used as sexual objects and cast aside once their utility has been exhausted, this is not definitive of sexual desire. By awakening us to the living presence of someone else, sexuality can enable us to treat this other being as just the person he or she happens to be. There is nothing in the nature of sexuality as such that necessarily reduces persons to things. On the contrary, sex may be seen as an instinctual agency by which persons respond to one another through their bodies" (Singer 1984, p. 382).

The spin-off of these developments and ideas is a much more relaxed and tolerant attitude toward sex in general in our current-day culture, although there are countries in the world (e.g., countries under the strong influence of traditional Islam, such as Afghanistan, Iran, and Saudi Arabia) where traditional sexual mores still prevail. The "new morality" also had the effect that some activities which traditionally were regarded as "sexual perversions," such as homosexuality, are much more widely tolerated and accepted in contemporary society (and decriminalized in most countries in the West), whereas there is much more sympathy for the practice of prostitution, even though it has only been legalized in a few countries or states. It will be noted later on what developments some of the ongoing debates about aspects of human sexuality have undergone.

Conceptual Analysis

The term "sex" is a multifaceted term which refers to many things. When used on its own, it could refer to the fact that most living organisms have an engendered existence which plays a central role in organic procreation. One can talk of "the sex" of an organism, thereby referring to whether a particular specimen of a species is male or female. The expression to "have sex" normally refers to the act of having sexual intercourse or interaction, either with a member (or members) of the opposite sex (heterosexual sex) or with a member(s) of the same sex (homosexual sex, sometime also referred to as gay and/or lesbian sex). The unqualified use of the term "sex" (as in questions and sentences such as "What do you think about sex?" or "Sex is what they live for") refers to the widespread practice of sexual engagement and activities, ranging from hetero- and homosexual intercourse to masturbation and fetishism.

The rest of this section will be focused on a single conceptual issue that tends to obfuscate clarity when sex is the topic of conversation or reflection. This is the issue of what exactly is meant by a "sexual activity." This discussion will partly draw on an analysis originally offered by Alan Soble.

What exactly does it mean to "have sex" or to engage in activities that actually count as "sexual activities"? Penile-vaginal penetration might be an uncontroversial case in point. But what about "heavy petting" that might not entail such penetration but might include activities such as fondling

breasts or "deep kissing" or even mutual masturbation? Do all of these count as "having sex"? In 1999, Stephanie Sanders and June Reinisch published the results of a survey that they conducted among undergraduate college students to establish these young people's perceptions of which of these kind of activities actually constitute "having sex" in *JAMA* (*Journal of the American Medical Association*). The results were quite surprising. While there was broad consensus about the sexual nature of penile-vaginal penetration, it was particularly surprising to learn that "only 40% indicated that they would say that they had 'had sex' if oral-genital contact was the most intimate behavior in which they engaged (60 % would not)" (Sanders and Reinisch 1999, p. 276).

Soble (n.d.) interprets this finding sympathetically. He argues that it shows that the notion of "sexual activity" is, for many people, a quite technical notion, whereas "having sex" is an "ordinary language concept, which refers primarily to heterosexual intercourse." This can also be gleaned from the racist prohibition of "sex across the color line" that occurred in one of the most notorious discriminatory laws in apartheid South Africa, viz., Article 16 of the so-called Immorality Act (happily repealed in the mid-1980s). The question that this grossly racist law raised, among others, was "when is interaction between people an instance of 'having sex'"? (The issue of when/whether the "color line" was breached was fortuitously dealt with by recourse to the similarly notorious "Population Registration Act"). This did not deter apartheid's legislators. The act went on to describe in tasteless detail exactly when that eventuality obtains (nowadays people mockingly recall that this article in the law contained the only instance of legal pornography that, at that time, was available in South Africa).

Maybe the safest way to try and understand the notion of "sexual activity" is to claim that the notion is defined by any activity from which sexual pleasure is derived (Gray 1997, p. 61). This idea would cover "normal" acts of heterosexual and homosexual intercourse, as well as masturbation and fetishism. The problem with such a definition is that it would then exclude activities that are sexual in intention, but, to use a phrase of Soble (n.d.), are "unsuccessful in a non-moral sense." Take rape as an example. While the perpetrator of a rape might derive some sexual pleasure from it (there are strong theories that rape is not so much a sex act but an act of violence), the victim does not. Are we then justified to deny that rape is, at least from the perspective of the victim, a "sexual act"? Less controversial might be the example of sex between hetero- or homosexual partners of long standing who have actually lost sexual interest in each other and simply engage in intercourse routinely without deriving any pleasure from it. When Gray's definition is applied to this last case, the intercourse between these partners does not count as a "sexual act" because no pleasure is derived from it. The most we could say is that these individuals tried to engage in a sexual act but were unsuccessful. I agree, in conclusion, with Soble: "It may be a sad fact about our sexual world that we can engage in sexual activity and not derive any or much pleasure from it, but that fact should not give us reason to not call these unsatisfactory events 'sexual'" (Soble n.d.). A more successful definition of "sexual act" than that of Gray would therefore be a sexual act is an act from which sexual pleasure is normally derived with the consent of the participant(s) or that consists of efforts at or imitations of the latter, even if pleasure is not forthcoming from such efforts or imitations.

Sex and Ethics

This final section of the entry deals with broad developments that can be observed in the more recent history of moral reflection on human sexuality. It has already been shown that sex, to a significant extent, constituted a kind of prototype or most visual instance of human sin and corruption. This particularly comes to the fore in the debate about sexual perversion, i.e., the issue as to whether sex functions in its "proper place" in human society or whether it is perverted for purposes that it was not meant for originally. This debate also provides a neat demonstration of key aspects in the understanding of human sexuality that have undergone significant changes over recent times.

The ethics of St. Thomas Aquinas, the influential medieval philosopher of thirteenth-century scholasticism, is a key point of reference in this debate. His is one of the most impressive (and later on, most controversial) efforts to understand sexual perversion in terms of natural law. The natural purpose of sexual activity for Aquinas is human procreation. In this respect, sex serves the same purpose for humans that it serves for the rest of the animal kingdom. That procreation is the aim of sex and the origin of the sexual impulse can best be motivated by noting the shape and working of the genitals. The penis and the vagina are shaped in the form that they have in order to optimally facilitate the process by which male sperm is deposited in the body of the woman, where it (via processes quite unknown to Aquinas) develops into a fully fledged human being. Sperm may not be deposited anywhere else; if so, it would be "unnatural" in the sense of being irreconcilable with God's plan or design. It is interesting to note that Aquinas, unlike St. Augustine and the church fathers, does not have much of a problem with the fact that coitus is accompanied by pleasure, as long as the pleasurable act of coitus occurs only between husband and wife and for the sole purpose of bringing about new life.

Sexual perversion, in terms of this view, occurs whenever sexual activities violate this single purpose. Masturbation, homosexual sex, fellatio, anal sex, and the like – acts which result in sperm emissions outside of the vagina – are all regarded as perversions of sex and therefore morally wrong. Premarital sex and prostitution, where coitus in its seemingly "natural" form does occur, are nevertheless similarly viewed as perversions because they are acts occurring between people that are not married. What is important is that Aquinas derives his notion of "natural sex" entirely from the form and shape of the human (or animal) genitals. The nature of the psychological relationship that normally accompanies sexual relationships is seemingly of no relevance to him. He accepts what Soble (n.d.) calls a purely "anatomical" criterion of natural and perverted sex. Additionally, it should be noted that perverted sex for Aquinas is necessarily also immoral sex, since it forfeits that "natural order of the world" that God has created and intended for humanity.

The notion of "perverted" sex – even if its "perversion" is not necessarily regarded as morally wrong – reappears in the reflections of contemporary thinkers like Thomas Nagel (1979). What is interesting is that while St. Thomas' notion of sexual perversion stems from his observation of the common, universal, "natural" shape and function of the genitals that we share with the rest of the animals, Nagel develops a notion of sexual perversion that stems not from what we have in common with other animals but from an essential aspect in which we are different from these other creatures. This unique characteristic of humans has nothing to do with the nature or form of (any aspect of) their anatomy, but has everything to do with their ability to foster intersubjective relationships at the psychological level.

Nagel draws on ideas of Jean-Paul Sartre to develop his position. Sartre talks of sex as a "double reciprocal incarnation." This occurs in a caress in the following way: "I make myself flesh in order to impel the Other to realize for herself and for me her own flesh, and my caresses cause my flesh to be born for me in so far as it is for the Other flesh causing her to be born as flesh" (Sartre 1969, p. 391). Sexual interactions for Nagel are the result of a complex process in which two people (it need not be a man and a woman; they could be of the same sex) recognize (not necessarily simultaneously, but eventually in the course of a sexual encounter) that they each, individually, desire the other and that they each, individually, are the object of the other's bodily desire.

Nagel's argument consequently is that perverted sexual encounters or events would be those in which this mutual recognition of arousal is absent. These would be situations where one of the partners would remain either fully subject or fully object, and the "double reciprocal incarnation" would be absent. The pattern of arousal and sexual consciousness is "incomplete." What must be noted is that in this representation of sexual perversion, there is no reference to sex organs. It is also not important whether the sexual encounter has as its intention, or results in, human procreation. This implies that, for Nagel, it is not essential that the nature of this encounter be limited to heterosexual contact; this mutual arousal can

certainly occur in homosexual and lesbian relationships. Penile-vaginal coitus is also not necessary for this experience; fellatio, cunnilingus, and anal intercourse could as well express this experience. It does, however, imply that activities such as prostitution, fetishism, pedophilia, and necrophilia must be regarded as perversions in terms of Nagel's position.

We live in a time where the approach and sentiments of a position such as that of Nagel have largely won the day, over and against the metaphysical naturalism of Aquinas (although not everyone who defends a more liberal view of sexual morality will necessarily agree with all the details of Nagel's argument). The Roman Catholic Church (which has, at one point in its history, recognized the work of St. Thomas Aquinas as its "official philosophy") has a membership of 1.2 billion people all over the world, of which 40 % reside in South America. This church's position on heterosexual marriage as the exclusive condition for morally legitimate sexual activity, its insistence that sex is primarily meant for human procreation, and its consequent opposition to artificial measures of birth control such as condoms and oral contraceptives (as is argued in papal encyclicals such as Paul VI's *Humanae Vitae* in 1968) remain unchanged. However, it is unlikely that the majority of Catholic members adhere to these prescriptions.

The Catholic Church has, in addition, been severely criticized for its opposition to condom use, particularly in view of the devastating scope and consequences of the HIV/Aids pandemic for which condom use remains, to this day, generally accepted as the most reliable protection. (Abstinence, which the church recommends, is, of course, de facto more reliable, but it is unrealistic to expect that abstinence can be effectively maintained in hetero- and homosexual relationships, let alone casual, sexually charged encounters.) In most other churches the idea that heterosexual sex primarily serves as a source of relaxation and pleasure seeking in human relationships, and only secondarily as the mechanism for procreation, has largely won the day (few churches have the same position with reference to homosexual and lesbian sex). In the Muslim world strictly conservative views on legitimate sex also obtain, particularly as far as women are concerned. In Islam, polygamous marriages (i.e., situations wherein men can have several wives, but not the other way round) are, however, allowed. In general, there has been a trend, particularly in Western Europe, North America, and most countries that are liberal or social democracies toward secularization, with the effect that the hold of ecclesiastical or religious mores on people has tended to weaken or diminish in our time.

The normative view that is seemingly gaining significant ground in our time is the belief that, ideally, human sexuality is much more than the primal mechanism for human procreation. One of the hallmarks of the psychological makeup of human beings is the need for intimate relationships with other humans. These "others" may be of the opposite or the same sex, and the relationships could be monogamous or polygamous. (Choices between the latter are mostly culturally influenced or determined.) Sex is a pivotal aspect of these relationships because it provides or embodies the most pronounced expression of intimacy that humans are capable of.

It is widely accepted that responsible (interpreted to the point of declaring it legal) indulgence in sexual activities requires a certain level of psychological development or maturity. Hence, all cultures frown on the idea that children (defined as people under the age of 16 or sometimes 18) engage in sex, and there is global consensus on the need for significant social and legal penalties for anyone who commits pedophilia.

There are still reservations about homosexual or lesbian sex in most cultures, but these reservations are notably less pronounced or public. The view that sexual orientation is probably physiologically determined, and that even if it is not, adults still have the right to choose sexual partners on the basis of mutual consent, is winning the day all over the world. In a number of liberal or social democracies, homosexual and lesbian sex has therefore been legitimized, and in some countries same-sex marriages, or prenuptial contracts between same-sex partners that bestow similar protections on partners to those of heterosexual agreements or marriages, are nowadays allowed and recognized.

Conclusion

Although there is no necessary conceptual or moral link between sex and love, most people tend to link these notions in their understanding of erotic love and in their practice of sex. As previously mentioned, sex is the ultimate expression of intimacy in human relationships. Most people do not simply indulge in sexual activity with other people purely for the sake of attaining orgasm. The sexual interaction itself is experienced as the most private, but therefore also the most revealing and endearing expression of sustained commitment to another human being. If this was not the case, it is unclear why people so universally still insist on either marriage or long-term cohabitation and not simply revert to masturbation (i.e., sexual self-stimulation to the point of orgasm). Masturbation, in turn, was for a long time (and in some circles still is) a sexual taboo, although studies have shown that it is much more widely practiced than realized earlier. Yet, with obvious exceptions, masturbation, for most people, is not enough. Sex seems, for most people, to only attain its force as the deepest and most sincere expression of intimacy, properly or adequately experienced in the context of longer-term relationships between adults who prefer to engage in mutual commitments.

Cross-References

- ▶ Ethics
- ▶ Homosexuality
- ▶ Human Body
- ▶ Human Nature
- ▶ Values

References

Augustine. (2009). On marriage and concupiscence (Book I) (trans: Holmes, P., Wallis, R. E., & Warfield, B. B.). In P. Schaff & K. Knight (Eds.), *Nicene and post-Nicene fathers, first series* (Vol. 5). Retrieved from http://www.newadvent.org/fathers/15071.htm. Accessed 22 April 2015.

Freeman, C. (2002). *The closing of the western mind*. London: Pimlico.

Gray, R. (1997). Sex and sexual perversion. In A. Soble (Ed.), *The philosophy of sex* (3rd ed., pp. 57–66). Lanham: Rowman & Littlefield.

Kant, I. (1948). *Groundwork of the metaphysics of morals* (trans: Paton, A. J.). New York: HarperTorch Books. (Original work published 1785).

Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge & Kegan Paul.

Nagel, T. (1979). *Mortal questions*. Cambridge: Cambridge University Press.

Plato. (1951). *The symposium* (trans: Hamilton, W.). Harmondsworth: Penguin Books.

Plato. (1961). Timaeus (trans: Jowett, B.). In E Hamilton & H Cairns (Eds.), *Plato: The collected dialogues* (pp. 1151–1211). Princeton: Princeton University Press.

Sanders, S. A., & Reinisch, J. M. (1999). Would you say you "had sex" if. . .? *JAMA, 281*(3), 275–277. doi:10.1001/jama.281.3.275.

Sartre, J. P. (1969). *Being and nothingness* (trans: Barnes, H. E.). London: Methuen. (Original work published 1943).

Scruton, R. (1986). *Sexual desire: A philosophical investigation*. London: Phoenix Press.

Singer, I. (1984). *The nature of love. Vol. 2: Courtly and romantic*. Chicago: University of Chicago Press.

Soble, A. (n.d.). Philosophy of sexuality. In *The internet encyclopedia of philosophy: A peer-reviewed academic resource.* Retrieved from http://www.iep.utm.edu/sexualit/print. Accessed 22 April 2015.

Further Readings

Baker, R. B., Wininger, K. J., & Elliston, F. A. (Eds.). (1998). *Philosophy and sex* (3rd ed.). New York: Prometheus Books.

Bloom, A. (1993). *Love & friendship*. New York: Simon & Schuster.

Primoratz, I. (1999). *Ethics and sex*. London: Routledge.

Soble, A. (Ed.). (1980). *The philosophy of sex: Contemporary readings*. New Jersey: Rowman and Littlefield.

Slippery Slope

Douglas Walton
Centre for Research on Reasoning, Argumentation and Rhetoric, University of Windsor, Windsor, ON, Canada

Abstract

In global bioethics, the slippery slope argument has been used in such issues as abortion, legalizing marijuana, physicians having to disclose their HIV status, euthanasia, and gene therapy. Its uses

in ethical controversies on the latter two topics have been most prominent, and the main examples treated in this entry are slippery slope arguments about euthanasia and gene therapy.

Keywords

Ethical argumentation; Dam burst argument; Thin edge of the wedge; Vagueness; Eugenics; Euthanasia; Deliberation

Introduction

The slippery slope argument goes by different names such as the thin edge of the wedge argument, the dam burst argument, the Pandora's box argument, the domino argument, the snowball argument, and the camel's nose in the tent argument. But depending on how you define these other arguments, they may be different from the slippery slope argument. Because the slippery slope argument is both highly intuitive as a form of argument and highly complex in its logical structure, it has resisted attempts to provide a precise and comprehensive definition. Slippery slope arguments are often mixed up with related arguments, such as argument from negative consequences, that are inherently different from the slippery slope argument. Another problem is that slippery slope arguments are typically put forward in a compressed way that conceals implicit premises drawn from common knowledge. This means that identifying a slippery slope argument in ethical controversies has been a difficult or even an impossible problem. Although different types of slippery slope arguments have been identified in the literature (Govier 1982; Walton 1992), so far no central type of slippery slope argument that they fall under has been identified. In this entry, four identifying characteristics are formulated for determining whether an argument is a slippery slope argument or not.

Yet another problem is that although slippery slope arguments have often been assumed to be fallacious by the logic textbooks, many in the literature accept the hypothesis that they can sometimes be reasonable. This state of affairs raises the question of how slippery slope arguments can be evaluated so that one can tell which are the reasonable ones and which are the fallacious or defective ones. This short entry cannot solve these problems, but can only give the reader some idea of what slippery slope arguments are like in the examples treated. It will be shown that a slippery slope argument has four basic components: (1) the first step, an action or policy being considered, (2) a sequence in which this action leads to other actions, (3) a so-called gray zone or area of indeterminacy along the sequence where the agent loses control, and (4) the catastrophic outcome at the very end of the sequence. The idea is that as soon as the agent in question takes the first step, he will be impelled forward through the sequence, losing control so that in the end he will reach the catastrophic outcome. Not all of these components are typically made explicit, however, in examples of the slippery slope argument seen in biomedical ethics.

Historical Background

The slippery slope argument is closely related to the *sorites* (heap) paradox known to the ancient Greek philosophers and attributed to Eubulides (Kneale and Kneale 1962, p. 114). The heap paradox can be formulated as an argument with two true premises leading by an apparently valid argument to a false conclusion.

Premise 1: If you take one grain away from a heap, it makes no significant difference – you still have a heap.

Premise 2: Each time you repeat this step, it makes no difference, because one grain is too small to make a difference between something being a heap or not.

Conclusion: Even if you only have a few grains left, after repeating this step in premise 2 many times, what you have left has to be a heap.

This argument represents a paradox because if a deductive argument is valid and the premises are all true, then the conclusion cannot be false.

This form of argument is similar in general outline to the slippery slope argument because the problem in both instances is related to vagueness, specifically the gray area on a continuum where a specific criterion cannot be applied to differentiate one thing from another. Sometimes the heap paradox is also called the bald man paradox. Consider a man who starts pulling out the hairs on his head one at a time. When he starts this procedure, it is clear that he is not bald, but at some undefinable point, it has become clear that he is bald. Sometimes the same form of argument or a comparable one is also called the argument of the beard, for the same reasons. At what point does a man go from clean-shaven to having a beard? The argument of the beard has been described in logic textbooks as the fallacy of arguing that there is no difference between two things because there is a continuum between them and no sharp dividing line between them (Thouless 1930). For example, in using this argument, one could contend that there is no difference between being rich and not being rich.

The heap paradox and the beard argument are different from the slippery slope argument, however, even though the latter two also evolve from vagueness. The slippery slope argument is about a decision to take action or adopt a proposal. It is an argument against an action because that action represents a first step in a sequence of actions on a continuum with a gray zone wherein the possibility of stopping cannot be pinpointed and where the sequence past that point inevitably (or with high probability) leads to a highly undesirable outcome. Vagueness here is part of the problem, but not the whole story.

It would appear that there is no direct historical link between the heap paradox and the slippery slope argument. The latter began to appear in textbooks on informal logic, for example (Beardsley 1966, p. 176), and has continued to do so. Although the slippery slope argument is typically included in such textbooks under the heading of informal fallacies, the textbook accounts do not generally claim that this type of argument is always fallacious and sometimes specifically state that is not. In recent history, the kinds of examples of the slippery slope argument that have been discussed extensively occur in law and biomedical ethics.

The First Example: Euthanasia

An example of the use of the slippery slope argument attributed to Bishop Sullivan has been much quoted in the biomedical ethics literature (Rachels 1986, p. 171).

> If voluntary euthanasia were legalized, there is good reason to believe that at a later date another bill for compulsory euthanasia would be legalized. Once the respect for human life is so low that an innocent person may be killed directly even at his own request, compulsory euthanasia will necessarily be very near. This could lead easily to killing all incurable cancer patients, the aged who are a public care, wounded soldiers, all deformed children, the mentally afflicted, and so on. Before long the danger would be at the door of every citizen. Once a man is permitted on his own authority to kill an innocent person directly, there is no way of stopping the advancement of that wedge.

In this example some elements of the slippery slope argument can be seen. There is the first step where voluntary euthanasia is legalized. There appears to be a sequence of events flowing from this initial event: respect for human life is less, which leads to innocent persons being killed directly, which takes us to compulsory euthanasia, which takes us to killing incurable cancer patients, the aged, wounded soldiers, deformed children, mentally afflicted persons, and so forth. It is not clear, however, whether all these events are supposed to happen at more or less the same time or whether there is some sequence whereby one leads to another. However, compulsory euthanasia is supposedly being taken as a very bad policy, which includes outcomes such as killing deformed children and so forth. So there do appear to be some of the elements of the slippery slope argument here, assuming that having compulsory euthanasia is the catastrophic outcome at the end of the slope. Finally, there is the statement at the very end that there is no way of stopping the advancement of the sequence.

It is this last statement which brings up consideration of whether the slippery slope argument can properly be evaluated as a fallacy in this case. For the argument is a prediction about what

might or will happen in the future, and such an argument is always a contingent matter. Even if voluntary euthanasia were to be legalized, it might well be possible that some bright line, some legal guideline for distinguishing between compulsory euthanasia and voluntary euthanasia, could be devised that would be found to be workable and adequate to prevent compulsory euthanasia from being permitted.

According to Saliger (2007, p. 342), slippery slope arguments are called dam burst arguments in the biomedical ethics literature in Germany. But these appear to be two different kinds of arguments. In a slippery slope argument, the agent is proceeding gradually down a slope but can still turn back until he gets to a gray area of indeterminacy where he loses control. In the dam burst argument, once the dam bursts, the water floods out flooding the area below. The flooding procedure may happen gradually, but there is no element of a sequence containing a gray area where the agent loses control over its actions. For these reasons, Rachels' example may be better classified as a dam burst argument rather than a slippery slope argument. It is impossible to tell until there exists a set of identifying characteristics to define a slippery slope argument.

Defining the Slippery Slope Type of Argument

A genuine slippery slope argument should have four basic characteristics. First, there is a framework of discussion in which two agents, in the simplest case, are deliberating on whether to take an action or policy that they are considering or that one of them is considering. Let us call the agent who is considering taking the action the proponent and the other party who is raising some doubts about whether the action is prudent the critic. Second, the critic postulates a sequence of actions that will flow progressively from the first action being considered. Third, although the first steps in the sequence may be harmless, the proponent will gradually be propelled along the sequence to actions that are progressively more serious. Fourth, at some indeterminate point that cannot be defined in advance, the so-called gray area, the agent loses control and can no longer stop the sequence of actions from moving ahead to the final catastrophic outcome. This set of identifying characteristics is based on the existence of a continuum from a first step to a final outcome, and the continuum is the basis of an argumentation model (Walton 1992). This model has been shown to apply to the kind of slippery slope argument used in the ethical issue surrounding genetic therapy by Launis (2002, p. 174). This example is outlined below.

Using these characteristics, a slippery slope argument is defined as one in which the agent initially has control over its actions and can still stop the descent toward the ultimate catastrophic outcome but at some point loses control so that after that point the catastrophic outcome has become inevitable. So, although the slippery slope argument is a prediction about what will happen in the future, and it is generally an exaggeration to claim that any prediction of the sort has an outcome that is inevitable, there is an aspect of inevitability involved. The tricky part of the argument is that at some indeterminate point along the sequence of actions, the agent loses control. In other words, after the agent has proceeded through this gray zone, then the final outcome may be described as inevitable. But this is tricky because if you claim that the final outcome is inevitable from the first step, that claim would make the slippery slope argument fallacious, because a prediction so far into the future cannot claim to be certain. For this reason, considerable care needs to be taken in formulating such an argument.

Reasonable Slippery Slope Arguments

A simple example of a reasonable slippery slope argument is a good starting point to discuss slippery slopes. Let's consider the hypothetical case of a father warning his son not to take drugs, even if one of his schoolmates offers a free sample and tells him that taking the drug will make him feel good and is harmless. The father advises that although taking this drug will make him feel

good, it is not only dangerous but can lead to a situation of dependency where the body craves more and more of the substance. As the drug continues to be taken to achieve the same effect, it will get harder and harder to stop taking it, and at some point it will become impossible to stop taking it because of the withdrawal symptoms, such as cravings and nausea. He advises his son that he will become addicted to the drug. He tells him that because continuing to take the drug has side effects that are harmful to health, the end result is that he will become sick, will not be able to carry on with his daily activities, and his life will be ruined. He tells his son that even with treatment, it may be difficult or impossible to stop taking the drug.

This example can be classified as a slippery slope argument because it has all four characteristics described above. But nevertheless it is hard to deny that it is an extremely reasonable argument for the father to use to try to persuade his son not to take drugs. Of course some drugs can be more addictive than others, and it will complicate the example by naming a specific drug, such as cocaine or heroin. But generally it can be said that this sort of example is in principle a reasonable kind of argument.

Other non-fallacious examples of the slippery slope argument can be found in legal arguments. A series of cases on the flag-burning issue ruled on by American courts can be found in Walton (1992, chapter 7).

The Example of Genetic Therapy

Somatic gene therapy refers to the insertion of therapeutic genes into the somatic cells (the nonreproductive cells) of a patient. In somatic gene therapy, the effects are restricted and not inherited by the person's offspring or later generations. Germ line gene therapy carries its effects over to future generations (Launis 2002, p. 170). Somatic gene therapy is currently used to treat genetic disorders such as immunodeficiencies, hemophilia, and cystic fibrosis and is also used to avoid complications of organ rejection by inserting bone marrow (Resnik 1994). Some countries such as Australia, Canada, Germany, Israel, Switzerland, and the Netherlands prohibit treating patients with germ line gene therapy, even though it shows promise for treating some genetic disorders.

Recently there is some apprehension that the introduction of germ line therapy is the first step in a slippery slope that will ultimately force the adoption of genetic enhancement. Gardner (1995) has described a series of steps that will move us further along the slippery slope sequence that will result in standardized use of germ line therapy as a medical reproductive service available to parents to genetically improve the cognitive abilities of their children. He describes several forces that will move this sequence of steps forward. One is technological development of the kind that always tends to drive us forward to use any new technology that promises to be successful. Another is that as genetic therapy becomes more and more successful, it will achieve greater public acceptance (Holtug 1993, p. 414). Once parents start using it, they will see the advantages of it for their children. Once it becomes more widely adopted, it will become apparent that the children who have benefited from genetic enhancement technology will do better in competitions, in test scores, and in school grades. Once this stage is reached, genetic enhancement will begin to be adopted by nations because they are constantly competing to promote higher rates of economic growth, so they will take a strong interest in producing skilled and well-educated children who are joining the workforce. As nations compete with each other, they will see genetic enhancement as a way to compete with other nations to achieve their economic growth.

Gardner does not see this sequence of events as inevitably leading to the adoption of genetic enhancement all over the world; he only sees it as a highly probable outcome of events that will be driven forward by natural forces, once the first steps are taken to accept germ line therapy. There is also another respect in which the argument he describes is different from the slippery slope argument. He does not see genetic cognitive enhancement as a bad thing, something with a negative value, or as a catastrophic outcome that should be avoided at all costs.

However, others have taken this argument another step forward by describing the outcome as the acceptance of eugenics. For historical reasons, this term has emotional connotations for many and might be perceived as a scary outcome that is fearful and dangerous. The term eugenics, coined by Francis Galton in 1884, refers to a social program for promoting the higher reproduction of people with valuable characteristics and the reduction of the reproduction of people with characteristics deemed to be undesirable (Bashford and Levine 2010). Eugenics was popular in America in the nineteenth century and early twentieth century but became discredited when it was used as a justification for the racial policies of Nazi Germany. Because of this historical association, the word eugenics has highly negative connotations, suggesting or implying something fearful and even evil.

Cognitive enhancement of children's abilities does not sound like such a bad thing in itself. It could even be seen as a good thing, perhaps increasingly so as genetic technology improves, for the reasons adduced by Gardner. However, as soon as genetic cognitive enhancement is linked to the word "eugenics," people definitely take a step back from it, and it might well be seen as a catastrophic outcome of a kind to rightly be fearful about. So once genetic cognitive enhancement is associated with eugenics, taking the step of moving forward with accepting germ line therapy could be setting a slippery slope sequence into place. But even without eugenics coming into it, germ line therapy can be attacked using value-based argumentation on the grounds that it could worsen existing economic and social inequities (Resnik 1994, p. 32).

Compressed Slippery Slope Arguments

Slippery slope arguments are typically put forward in natural language argumentation in a compressed form in which parts of the argument indicated by the four characteristics of the slippery slope argument given above are not explicitly stated. Based on common knowledge, these components are only present in an enthymematic (non-explicit) form. An enthymeme, in logic, is an argument that has some of its components implicitly indicated, but not explicitly expressed in the text of discourse in which the argument was put forward. Usually an argument is said to be an enthymeme if it has a missing premise, but sometimes the not explicitly stated component can be the conclusion. The classic example is the following argument: all men are mortal; therefore, Socrates is mortal. The unstated premise is the proposition that Socrates is a man. This premise can be taken to be generally known in common knowledge, and since its insertion into the original incomplete argument would make the argument valid, there are grounds for taking the argument to be also based on this implicit premise along with the explicit premise that all men are mortal.

Slippery slope arguments often have this feature of being incompletely expressed or enthymematic. Consider the following example: if voluntary euthanasia is legalized, in the future there will be more cases of medical murder. Because of common knowledge, an audience can understand that there is some sort of implied transition so that there are a series of steps between the first action of legalizing voluntary euthanasia and the undesirable outcome of more cases of medical murder. Murder is something very bad; indeed it is a crime, a so-called capital offense. Hence, this outcome can be taken to have a high negative value, something that one would definitely want to avoid because it is dangerous. But what the steps are between legalizing voluntary euthanasia and the outcome of more cases of medical murder is left entirely implicit. Even so, if it can be assumed that there is an intervening sequence of events between these two polar events that link the polar events together going over some gray zone in which there is a loss of control leading down the slope to the catastrophic outcome, then the initial argument from negative consequences can be identified as a slippery slope argument. Such a classification rests, however, on several implicit assumptions that are only thinly supported at present by the original formulation of the argument. In such cases, the critic needs to test the original argument as formulated and try to get its proponent to fill in the missing steps. Otherwise, there is danger of confusion and

misdirection, and this is the kind of tricky case associated with fallacious uses of the slippery slope argument.

Close Relationships to Associated Types of Arguments

The slippery slope argument is related to several other forms of argument that are not identical to it but are often confused with it. Argument from negative consequences cites the consequences of a proposed course of action as a reason for not taking that course of action. Argument from negative consequences has the following general form (Walton et al. 2008, p. 332):

Premise: If action *A* is brought about, negative consequences will plausibly occur.
Conclusion: Therefore, *A* should not be brought about.

This structure represents a kind of reasoning that is used all the time in natural language deliberations. It is a reasonable argument, and the slippery slope argument is a subspecies of argument from negative consequences. But argument from negative consequences is not the same kind of argument as slippery slope argument. It is more general. In order to be a slippery slope argument, a given argument has to fit the scheme for argument from negative consequences above but it also has to have several other components. The action *A* has to be the first step in a sequence leading through a gray zone to an ultimate outcome that has an unusually high negative value. Moreover, there also has to be an element of loss of control involved that takes place during the gray zone.

Two examples from Corner et al. (2011, p. 135) help us to distinguish between slippery slope arguments and arguments from negative consequences. One example they cite is the argument opposing the legalization of cannabis because it would lead to an increase in lung disease. This is an example of argument from negative consequences, but as the example is stated, it cannot be classified as an instance of the slippery slope argument because the evidence that it has the other required characteristics of the slippery slope argument is not there in the example. An example of an argument they consider as fitting the requirements of a slippery slope argument is this one: if cannabis were legalized, attitudes toward harder drugs might become more positive, and in the future heroin might also become legalized. This argument can be classified as a compressed version of the slippery slope argument given that the final outcome of legalization of heroin is being put forward as an outcome with high negative value and given that there is an implicit sequence from the first step of legalization of cannabis to other steps in which progressively harder drugs are legalized, then the sequence is driven forward by an increase in positive attitudes toward harder drugs. These intervening steps are not stated explicitly, which is a defect of the argument if it is to be classified as a proper slippery slope argument. Nevertheless enough of the characteristics of the slippery slope argument are present for provisionally classifying this argument as fitting the requirements for a slippery slope argument.

The slippery slope argument is also closely related to a type of argument that is familiar in logic called *reductio ad absurdum*. In this type of argument, a hypothesis is put forward and an absurd consequence is drawn from it by a sequence of logical reasoning. Typically the absurd consequence is a logical inconsistency. Since a logical inconsistency is a false statement, by the rule of deductive inference called *modus tollens*, the hypothesis itself must be false. The rule of *modus tollens* has the following form in deductive logic, where *P* and *Q* are propositions:

If *P*, then *Q*.
Not *Q*.
Therefore, not *P*.

Once again though, the point needs to be made that the slippery slope argument is not identical to the form of argument called *reductio ad absurdum* but is a subtype of it. The slippery slope type of argument also takes values into account and is therefore related to a simpler form of argument called argument from values.

Argument from negative consequences is based on the assumption that consequences of an action can be designated as having a negative value, and the slippery slope argument also has this feature when it postulates the ultimate outcome of the slope will be catastrophic, something with a very high negative value. But arguments from positive and negative value are independent forms of argument in their own right that give practical reasons for carrying out (or not carrying out) a contemplated action (Walton et al. 2008). Argument from negative value has the following general form:

Major premise: If action *A* has negative value *V*, *A* should not be carried out.
Minor premise: Action *A* has negative value *V*.
Conclusion: *A* should not be carried out.

It is important to note that this form of argument is defeasible, meaning that it offers a presumption in favor of its conclusion if the premises are accepted, but this presumption can be defeated by new evidence that comes into a case. So the finding that action *A* has negative value *V* which is a reason for not carrying out *A* is not final; it may be possible in a given case that there are countervailing reasons for carrying out *A*, for example, the finding that *A* has positive value greater than its negative value.

Domino Arguments and Dam Burst Arguments

The terminology used to identify and classify slippery slope arguments has not yet stabilized. But it can be helpful to propose that there are two types of arguments often equated with the slippery slope argument that need to be seen as special instances of it but are not equivalent to it as forms of argument. The first of these is the so-called domino argument. The domino argument has a sequence of events in which each one in the sequence causes the next one to happen in such a manner that once the first event occurs, it will lead to the next event, and so forth, until the last event in the sequence finally occurs. In slippery slope arguments, the sequence on which these actions are propelled forward is often at least partly causal in nature. Therefore, it seems reasonable to propose that on the definition of the slippery slope argument given above, the domino argument represents one part of the slippery slope argument. However, the domino argument does not postulate the catastrophic outcome as the last step in the sequence. It merely says that once the first step is taken, it will lead to a sequence of outcomes that will lead to loss of control and therefore might be more serious than was originally anticipated. Another difference between the domino argument and the slippery slope argument is that the domino argument requires no step by step participation by the decision-maker in the process of sliding down the slope (Saliger 2007, p. 343). How to define these terms is not yet settled, but if the domino argument is defined in this way, it is clearly different from the slippery slope argument but can be seen as a part of it and closely related to it.

The other type of argument often equated with the slippery slope argument is called the dam burst argument (Saliger 2007). In the case of the latter type of argument, there is one cataclysmic event, the dam bursting, and then there is either slowly or immediately the catastrophic outcome of the flooding of the area below the dam, presumably with loss of property and lives of those affected by the flooding. This metaphor suggests a special kind of argument from negative consequences that has a highly negative outcome, a ruinous disaster. In this respect, it is comparable to a slippery slope argument. But the metaphor of the dam bursting carries with it no essential element of a sequence of steps from an initial action through a gray zone with its accompanying loss of control eventuated in the ultimate outcome of the ruinous disaster. For these reasons, it seems best to propose drawing a distinction between dam burst arguments and slippery slope arguments.

It is also very useful for practical purposes in biomedical ethics to distinguish between the heap paradox known to the Greeks and the slippery slope argument. Clearly the slippery slope argument is related to the paradox because an essential

component in both structures is the gray zone of indeterminacy within the sequence of smaller steps. And it is true that in many instances, especially in legal arguments, the slippery slope is due to the vagueness of the key term that is hard or tricky to define precisely. While it is very useful to recognize that what makes a slippery slope argument slippery is vagueness (the gray zone), the slippery slope argument has other key identifying characteristics connected to its vagueness that makes it distinctive.

Conclusion

Slippery slope arguments can be reasonable as suggested by the example in which the father advises his son that taking drugs would not be a good choice of action because in the end it might easily lead to a catastrophic outcome. A reasonable slippery slope argument has four basic characteristics that can be summed up as follows: first, there must be a framework of deliberation in which one agent is advising another on a choice of action; second, there must be a sequence of actions from a first one to an ultimate outcome that is catastrophic; third, there must be a gray area during which the agent will lose control; and fourth, once in this area, the agent will inevitably be dragged forward toward the catastrophic outcome.

More extreme examples of the slippery slope argument of the kind classified as fallacious by the logic texts are cases where the argument is poorly supported by the evidence and in place of this uses exaggeration and fear in an emotional appeal. In other cases, the argument is not fallacious but is too weak, either because the premises are not supported by enough evidence or because the argument is put forward in a compressed form that leaves out required premises. In still other cases, the argument is open to available counterarguments that it has not rebutted or even considered.

Consider the slippery slope argument on euthanasia as an example. Using case-based reasoning, such an argument needs to be evaluated in an evidential situation where other arguments supporting or attacking it need to be taken into account. For example, a counterargument to a slippery slope argument against allowing euthanasia might be the argument that a system of counseling and psychiatric assessment can be put in place using physicians to judge whether the patient is making a voluntary decision. In such a case, the counterargument offers a mechanism whereby the indeterminacy of the gray zone can be offset by some rational criterion that enables a reasonable way of making decisions about the borderline cases. With this kind of counterargument in place, the slippery slope argument is defeated by a counterargument that attacks one of its premises. Unless the proponent of the original slippery slope argument can attack the attack with another counterargument, his argument is defeated. The general method whereby such defeasible arguments are evaluated is on the basis of burden of proof as it shifts back and forth in a larger network of argumentation. To evaluate ethical arguments by fairly considering both sides, it is necessary to construct an argumentation tree structure that models the pro and con arguments, taking the logical form of each argument into account.

Cross-References

- ▶ Bioengineering
- ▶ Euthanasia: Active
- ▶ Euthanasia: International Debate
- ▶ Future Generations
- ▶ Genetically Modified Organisms (GMOs): Human Beings

References

Bashford, A., & Levine, P. (2010). Introduction: Eugenics and the modern world. In A. Bashford & P. Levine (Eds.), *The Oxford handbook of the history of eugenics* (pp. 3–24). Oxford: Oxford University Press.

Beardsley, M. C. (1966). *Thinking straight*. Englewood Cliffs: Prentice-Hall.

Corner, A., Hahn, U., & Oaksford, M. (2011). The psychological mechanism of the slippery slope argument. *Journal of Memory and Language, 64*, 133–152.

Gardner, W. (1995). Can human genetic enhancement be prohibited? *The Journal of Medicine and Philosophy, 20*, 65–84.
Govier, T. (1982). What's wrong with slippery slope arguments? *Canadian Journal of Philosophy, 12*, 303–316.
Holtug, N. (1993). Human gene therapy: Down the slippery slope. *Bioethics, 7*(5), 402–419.
Kneale, W., & Kneale, M. (1962). *The development of logic*. Oxford: Oxford University Press.
Launis, V. (2002). Human gene therapy and the slippery slope argument. *Medicine, Health Care and Philosophy, 5*(2), 169–179.
Rachels, J. (1986). *The end of life*. Oxford: Oxford University Press.
Resnik, D. (1994). Debunking the slippery slope argument against human germ-line gene therapy. *The Journal of Medicine and Philosophy, 19*(1), 23–40.
Saliger, F. (2007). The dam burst and slippery slope argument in medical law and medical ethics. *Zeitschrift fur Internationale Strafrechtsdogmatik, 9*, 341–352.
Thouless, R. H. (1930). *Straight and crooked thinking*. London: English Universities Press.
Walton, D. (1992). *Slippery slope arguments*. Oxford: Oxford University Press.
Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. Cambridge: Cambridge University Press.

Further Readings

Schauer, F. (1985). Slippery slopes. *Harvard Law Review, 99*(2), 361–383.
Van der Burg, W. (1991). The slippery slope argument. *Ethics, 102*, 42–65.
Volokh, E. (2002). The mechanisms of the slippery slope. *Harvard Law Review, 116*(4), 1026–1137.

Social Ethics

Yesim Isil Ulman
School of Medicine, Department of History of Medicine and Ethics, Acibadem University, Atasehir-Istanbul, Turkey

Abstract

The ethical values and behaviors are not only abstract terms, but they are refined and conceptualized by real-life experiences. The societal context where the actions of humans can be analyzed by ethical decision-making is entirely relevant to deliberate on what is the right thing to do and what the moral agent should do, since the ethical values and principles response to the actual practices of life and to the needs of humans in the society. This elaboration takes us to the realm of social ethics.

This article reviews the definition and contextual meaning of social ethics at a broader level by giving special emphasis to the ethical theories and principles, focusing on the societal and public setting. Ethics will be deliberated with social and community aspects. Based on the principle of justice and public health ethics, the concept of social ethics has been investigated concisely through the relationship between man, as a moral person, and the society in exemplification of the issues of healthcare ethics. It is argued that the tension between individualism and communitarian needs can be reconciled with the perspective of social ethics by respecting the individual autonomy without disregarding the common good and social justice. By promoting the values of social responsibility, solidarity, and social utility, social ethics has been proposed as the basis of a rational, moral, egalitarian, pluralistic, democratic society rising on the pillars of human rights and human dignity.

Keywords

Ethics; Ethical theories; Ethical principles; Beneficence; Non-maleficence; Autonomy; Justice; Common good; Distributive justice; Equality; Individual freedom; Public health ethics; Public reason; Social rights; Social utility; Solidarity; Social responsibility

Introduction

Despite the fact that it has not been adequately defined in the contemporary bioethics literature, the concept of social ethics contains manifold denotations. Social ethics is inevitably fraught with the ethical theories and principles, the principle of justice in particular, as well as the moral codes of conduct and ethics of public health. Rather than having a precise definition, the concept of social ethics is in need of being employed with its derivatives such as "social" and "society" (Armstrong 1907).

The interaction between human (as a moral person) and society attests the evolution of mankind. In the beginning, man was forced to accept some social obligations in order to remain within a social group or clan which provided survival security, since life outside the clan was not viable.

Greek philosophy brought a new understanding of man: it underlined the value of the individual but acknowledged the conflict created between his (her) need to belong to a group and need to be recognized as an individual with specific characteristics. Aristotle, in particular, stated that the essence of a human being is not only his rationality but also his capability of relating to others, since man is naturally meant to live in a community.

Aristotle's ethical methodology lies on "good action" and defends that we study ethics in order to improve our lives, and therefore its principal concern is the nature of human well-being. Aristotle follows Socrates and Plato in taking the virtues to be central to a well-lived life. Like Plato, he regards the ethical virtues (justice, courage, temperance, and so on) as complex rational, emotional, and social skills. What we need, in order to live well, is a proper appreciation of the way in which such goods as friendship, pleasure, virtue, honor, and wealth fit together as a whole. In order to apply that general understanding to particular cases, we must acquire, through proper upbringing and habits, the ability to see, on each occasion, which course of action is best supported by reasons. Therefore practical wisdom, as he conceives it, cannot be acquired solely by learning general rules. We must also acquire, through practice, those deliberative, emotional, and social skills that enable us to put our general understanding of well-being into practice in ways that are suitable to each occasion (Craut 2014). In fact the idea and inspirations of social ethics may be rooted back in the Greek philosophy and Aristotle's ethics.

Thus, Aristotle's notions of political wisdom (*phronesis*) and political action (*praxis*) have been benefited to connect between the Greek ethical thought and the social ethics of Karl Marx, later on. Those elements in Aristotle's thought include emphasis on the social and political nature of man, the metaphysical relationship between substance and form, potency and act, and finally his distinctions among theoretical, practical, and productive knowledge. Divergent philosophical positions regarding social and economic justice – whether utilitarianism, formalism, liberalism, etc. – stress the distribution of social wealth in terms of individual happiness, social contract, or fairness. However, Marx believes that the essential questions of ethics and politics lie in the analysis of the nature and structure of the economic-social infrastructure and the organization of the productive relations (McCarthy 1986).

Since then, a balance has been sought between these two characteristics. Unfortunately, history has shown how this tension has brought about very negative consequences: in the name of the common good, some atrocities have been committed against individual freedom and dignity; in the name of individual rights, society has been considered by some individuals as a system that frustrates fulfillment of needs and expectations causing them to opt out. By considering how both excesses have their limits and their pitfalls, society is learning to value and respect the dignity of the individual, while the individual is discovering that society is not just for the reception of benefits but is the right place to fully express and fulfill himself as a human being in a state of interdependence (Di Mattia 2008).

Based on this brief historical perspective, social ethics can be interpreted through the definition, methodology, theories, and principles of ethics at a societal context, by giving special emphasis to the justice principle in ethics.

S

Ethics and Ethical Methodology

Ethics is defined, first of all, as standards of conduct that distinguish between right and wrong, good and bad, and so on; secondly ethics is an academic discipline that studies standards of conduct, and ethics is also described as a state of character. As an academic discipline, ethics is a branch of moral philosophy that is concerned with age-old questions about duty, honor, integrity, virtue, justice, and the good life. The questions

asked by moral philosophy are normative, rather than descriptive, in that they have to do in what one ought to live or how ought to be structured. Several disciplines in the social and behavioral sciences, such as psychology, sociology, anthropology, and political sciences, take a descriptive approach to ethical questions in that they attempt to describe and explain ethical beliefs, attitudes, and behaviors. The study of ethics can be subdivided into theoretical (or normative) ethics, which studies general theories, concepts, and principles of ethics; meta-ethics, which studies the meaning and justification of ethical words, concepts, and principles; and applied (or practical) ethics, which studies ethical questions that arise in specific situations or areas of conduct, such as medicine, research, and so on (Shamoo and Resnik 2009).

When ethics is defined as a standard of conduct, it is important to compare ethics with law. Societies have had laws since ancient times. One of the first legal systems was the 282 rules of the Code of Hammurabi (1795–1750 BC), established nearly 4,000 years ago. Modern legal systems are based, in large part, on laws developed in ancient Rome. Laws are like ethical standards in several ways. First, laws, like ethics, are standards of conduct: they tell people what they ought to and ought not to behave. Second, ethical and legal standards share many concepts and terms, such as duty, responsibility, negligence, rights, benefits, and harms. Third, the methods of reasoning used in law and ethics are quite similar: both disciplines give arguments and counterarguments, analyze concepts and principles, and discuss cases and examples. However ethics differ from law in several important ways as well. First, the scope of law ethics is much different from the scope of law. There are many types of conduct that might be considered unethical but are not illegal. Second, people can appeal to moral or ethical standards to evaluate or judge legal ones. People may decide there needs to be a law against some type of ethical behavior, or they may decide that an existing law is unethical. If we decide that a law is unethical, then we are morally obligated to change it or disobey it. For example, many people who decide that South Africa's system of apartheid to be unethical fought to change the system. Herewith, ethics and law can be imagined as circles that overlap in some areas (Shamoo and Resnik 2009). As the two disciplines fostering each other in terms of methodology and content, ethics and law can act together to respond to the needs of human welfare, common good, and social justice. The notion of social ethics is quite relevant and stimulating in this link.

In this context, another distinction is between ethics and religion. Ethical theories and religious traditions have much in common in that they prescribe standards of human conduct and provide some account of the meaning and value of life. However ethics is not the same as religion. First, people from different religious backgrounds can agree on some basic ethical principles and concepts. Christians, Jews, Muslims, Hindus, and Buddhists can all agree on the importance of honesty, integrity, justice, benevolence, respect for human life, and many other ethical values despite their theological disagreements. Second, the study of ethics or moral philosophy is a secular discipline that relies on human reasoning to analyze and interpret ethical concepts and principles (Shamoo and Resnik 2009). Consequently, the rational decision-making and secular way of thinking methodology can be implemented with the guiding principles of international human rights law and universal values in order to reach ethical analysis with social and community aspects.

Ethical Theories and Principles in Relation to Social Ethics

It is worth to examine the ethical theories in the sense of social and communitarian aspects. Beauchamp defines moral principles as respect for autonomy (the obligation to respect the decision-making capacities of autonomous persons), non-maleficence (the obligation to avoid causing harm), beneficence (obligations to provide benefits and to balance benefits against risks), and justice (obligations of fairness in the

distribution of benefits and risks). Method in ethics, firstly, begins with the moral convictions that inspire the highest confidence and that appear to have the lowest level of bias. They serve as first principles and conditions of more specific moral conceptions. Ethical principles are taken to be universally valid norms that warrant us in making intercultural and cross-cultural judgments about moral depravity, morally misguided beliefs, savage cruelty, and other moral failures. Secondly, these abstract principles need to be specified to make them suitable for the analysis of a context, case, or policy. It requires, as does the associated method of reflective equilibrium, that we match and adjust all of our well-substantiated moral judgments in order to render them coherent with the full range of our moral commitments (Beauchamp 2003). This view will be explored more in Rawlsian view of justice and public reason in terms of social ethics later in the paper.

Kantian ethical theory should be briefly dealt with in this respect. Kantianism is a theory developed by the German Enlightenment philosopher Immanuel Kant (1724–1804), which has been revised and fine-tuned by modern day Kantians, such as Christine Korsgaard (1996). The basic insight of Kantianism is that the ethical conduct is a matter of choosing to live one's life according to moral principles and rules. The concept of a moral agent plays a central theory in Kant's theory. A moral agent is someone who can distinguish between right and wrong and can legislate and obey moral laws. Moral agents (or persons) are autonomous (or self-governing) insofar as they can choose to live according to moral rules. For Kant, the motives of agents (or reasons of actions) matter a great deal. One should do the right action for the right reason. What is the right thing to do? According to Kant, the right thing to do is embodied in a principle known as the categorical imperative. One of the categorical imperatives holds that one should act in a way that one's conduct could become a universal law for all people. According to another Kantian categorical imperative, one should treat humanity always as an end, never as a means. The basic insight here is that human beings have inherent (or intrinsic) moral dignity or worth. We should not abuse, manipulate, harm, exploit, or deceive people in order to achieve specific goals (Shamoo and Resnik 2009). Thus Kantian ethical theory is also relevant in deliberating the moral agent's actions in terms of its societal implications.

Originated from the utilitarian moral philosophy of Jeremy Bentham (1748–1832) and John Stuart Mill (1806–1873), consequentialism (utilitarianism) bases ethical decision-making on an analysis of the likely consequences or outcomes of different choices and actions. A consequentialist is somebody who thinks that what determines the moral quality of an action (i.e., determines whether it is right or wrong) are its consequences. A contrast is sometimes drawn between theories which determine the moral quality of actions by their observance or nonobservance of rules and those which determine it by whether they promote valued consequences. It is, obviously, determined by both, and that any adequate theory will take both consequences and rules into account. We normally judge rightness or wrongness of actions by their conformity to rules or principles, and the principles themselves are judged by the consequences of observing them. If the actions are international, we praise or blame the agent for them (Hare 2012).

Beauchamp and Childress first outlined this contemporary theory of applied ethics in 1983. Their claim is that a decision is ethically sound provided certain principles are respected and balanced. They proposed four principles, although proponents have since suggested other candidates. Principlism has become one of the most popular theories in healthcare ethics, and the principles provide insights into ethical problem-solving (Schwartz et al. 2002). Ranaan Gillon argues that four principles have moral relevance in the application to healthcare ethics and predicts that they are going to be acceptable as the basis for a global bioethics, compatible with and sensitively negotiating the delicate path between moral relativism and moral imperialism and helping in the pursuit of morally acceptable world peace (Gillon 2003).

The most commonly applied principles are listed below.

Autonomy

The ability of a person to be self-determining and self-governing; the capacity of a person to make reasoned choices on the basis of information. It implies a duty on the part of caregivers to do what is necessary to promote or at least not hinder their patient's autonomy. This requires respect for persons, by not interfering with their plans, ambitions, and choices (recall Kant's categorical imperative regarding ends and means). Autonomy is the primary consideration in patient-centered treatment.

Beneficence/non-maleficence

These are related concepts. Beneficence requires the caregiver to do good and help people; non-maleficence is the Hippocratic requirement on the caregiver to do no harm. If only beneficence was required of a medical practitioner, it would be impossible to maintain because it entails no limits. Thus the requirement is balanced, so at the very least caregiver ought to do no harm. However, even this principle is not satisfactory on its own, as practitioners do occasionally have to cause some harm, such as the sting of a needle or a noxious treatment like chemotherapy. Thus we rely on beneficence to ensure that the harm is performed for a greater end.

Justice

In some of the literature justice means to treat people fairly. This might entail treating equals equally whenever possible. However, it might also mean treating some people differently when their differences are relevant. For instance, we might choose to provide more healthcare to low-income areas where health problems are often greater and healthcare is traditionally less accessible. Some philosophers believe justice means equality of distribution of resources, while others claim it requires only equality of access (Schwartz et al. 2002).

Every civilized society is a cooperative venture structured by moral, legal, and cultural principles that define the terms of social cooperation. Beneficence and respect for autonomy are principles in this fabric of social order, but justice has been the subject of more treatises on the terms of social cooperation than any other principle. A person has been treated justly if treated according to what is fair, due, or owed. For example, if equal political rights are due to all citizens, then justice is done when those rights are accorded (Beauchamp 2008). The terms fairness, desert (what is deserved), and entitlement have been used by various philosophers, in attempts to explicate justice. These accounts interpret justice (in medical setting) as fair, equitable, and appropriate treatment in light of what is due or owed to persons. Standards of justice are needed whenever persons are due benefits or burdens because of their particular properties and circumstances, such as being productive or having been harmed by another person's acts. An injustice involves a wrongful act or omission that denies people resources or protections to which they have a right (Beauchamp and Childress 2009).

Distributive justice and social ethics

The term distributive justice refers to fair, equitable, and appropriate distribution in society determined by justified norms of distribution that structure part of the terms of social cooperation. Usually this term refers to the distribution of primary social goods, such as economic goods and fundamental political rights. But burdens are so within its scope. Paying for forms of national health insurance is a distributed burden.

Recent literature on distributive justice has tended to focus on consideration of fair economic distribution, especially unjust distributions in the form of inequalities of income between different classes of persons and unfair tax burdens on certain classes. But many problems of distributive justice exist besides issues about income and wealth, including the issues raised in prominent contemporary debates over healthcare distribution (Beauchamp 2008).

Shift from Individualism Toward Rights and Duties in the Community

Robert Veatch comments that when modern ethics began to shift from a Hippocratic ethics of benefit to a more deontological ethics of rights and duties, drawing on the notion of respect for persons and the underlying principles of fidelity, autonomy, veracity, and avoiding killing, the new ethics was still addressing problems of the individual patient/physician relation – problems of confidentiality, informed consent, disclosure of diagnosis, and the care of dying patient. It was as if in all the world there were only one physician and one patient. The moral problem was figuring out how the patient ought to be treated. The dispute between the consequentialist Hippocratic ethics and non-consequentialist ethics of respect for persons was one within the tradition of individualism. Veatch argues that the moral problems in medicine of the future moved from individual to a more social model. This shift required confronting the problems of ethical individualism. Both Hippocratic beneficence and respect for persons ignore duties to third parties. In the modern world, ignoring society increasingly becomes impossible. Medicine must confront issues of allocating scarce medical resources, including organ transplant, and conducting research on human subjects where the goal is not improving the welfare of the individual patient but producing knowledge for the benefit of the society.

Veatch examines the concepts of social utility, resource allocation, equity, and justice to draw attention to the need of a social ethics. Social utility is defined as a principle that an action or rule is morally right insofar as it produces as much or more net good consequences as any alternative, taking into account the benefits and harms for all parties affected. To achieve this goal and to balance social utility and justice, egalitarian justice principle is implemented in the sense of allocating scarce resources justly on the basis of need and equity (Veatch 2003).

At this point, the ethics of public health is particularly relevant to improve the individualistic approach of ethics in favor of common good and to endorse the perspective of social ethics.

Public Health Ethics, as a Link Between Healthcare and Social Ethics

Public health is primarily concerned with the health of the entire population, rather than the health of individuals. Its features include an emphasis on the promotion of health and the prevention of disease and disability; the collection and use of epidemiological data, population surveillance, and other forms of empirical quantitative assessment; a recognition of the multidimensional nature of the determinants of health; and a focus on the complex interactions of many factors – biological, behavioral, social, and environmental – in developing effective interventions. Public health ethics, like the field of public health it addresses, traditionally has focused more on practice and particular cases than on theory. We can establish the relevance of a set of these considerations in part by looking at the kinds of moral appeals that public health agents make in deliberating about and justifying their actions as well as at debates about moral issues in public health. The relevant general moral considerations include:

- Producing benefits
- Avoiding, preventing, and removing harms
- Producing the maximal balance of benefits over harms and other costs (often called utility)
- Distributing benefits and burdens fairly (distributive justice) and ensuring public participation including the participation of affected parties (procedural justice)
- Respecting autonomous choices and actions, including liberty of action
- Protecting privacy and confidentiality
- Keeping promises and commitments
- Disclosing information as well as speaking honestly and truthfully (often grouped under transparency)
- Building and maintaining trust

Several of these general moral considerations – especially benefiting others, preventing and removing harms, and utility – provide a prima facie warrant for many activities in pursuit of the goal of public health. It is sufficient for our purposes to note that public

health activities have their grounding in general moral considerations and that public health identifies one major broad benefit that societies and governments ought to pursue (Childress et al. 2002).

Public health ethics, as a field of study, also, seeks to understand and clarify principles and values which influence decision-making in public health practice. Whereas public health ethics, as a field of practice, applies principles and values to public health activities, it helps to deal with ethical dilemmas, in order to come up with the best possible solution for a specific case. Public health ethics is also concerned with the ethical dimensions of public health as a specific profession (professional ethics).

Public health ethical issues were minimal when the paternalistic approach to public health measures was accepted (paternalism is the attempt to impose limitations upon someone or to require actions by someone for his or her own good; such impositions usually are justified with children because it is assumed that they are incapable of deciding on their own behalf and with those who, because of cognitive limitations, cannot choose on their own behalf). As from the second half of the twentieth century, the authority of medicine and the paternalistic approach of physicians and public health officials started to be questioned due to certain factors:

- The discovery of new treatments and new technologies has enormously increased the chances of cure along with the risk of causing harm.
- Patients' knowledge in general and on medical issues in particular has grown out of all proportion.
- The place of the individual has gained more ground in society, as have government policies and laws which have developed toward protecting the individual, being informed by autonomy and human rights issues.

Medical ethics has proportionally increased its body of study and research, in an effort to update its code to the new reality, which now includes problems like medicine and palliative medicine (palliative care and hospice). As a result, the more inclusive term for ethics in clinical medicine is "bioethics." Bioethics has found a strong point of reference in the patient-centered approach. As a result, clinicians have clear guidelines for their interaction with patients, based on four principles: beneficence (doing good), non-maleficence (not doing harm), autonomy (respect for personal rights and the individual), and justice (distributing benefits, risks, and costs fairly).

Those four principles have also strongly influenced public health decisions in the last few decades, based on the argument that, while directed to the whole population, public health activities interact with individuals. Although public health practice should not overlook the rights, interests, and freedom of the individual, it has to look at the well-being of the entire population; therefore, the ethical principles and values (ethical values) applied in bioethics, which follow the individualistic orientation, cannot be used as a point of reference when dealing with the entire population.

Let us take the principle of non-maleficence (not doing harm). If the point of reference is an individual, it can be (somehow) easy to identify a possible harm and, as a consequence, be easy to avoid that harm. However, if the case is of a necessary intervention targeted at a whole population, there may well be individuals likely to be harmed by the intervention who are difficult (if not impossible) to identify, and the harm, therefore, cannot be avoided. As a conclusion, in public health practice the principle of non-maleficence can be understood as doing the least harm possible to the least possible number of people (Di Mattia 2008).

Interdependence is the complement of autonomy; an individual has a social role which, if carried out, prevents the development of an extreme individualistic perspective that is inconsistent with the true nature of human beings. We all understand that a society where individuals are free to do whatever they want would not last long. Many of the collective goals in society that benefit the whole community are achieved by sacrificing some degree of independence and freedom; maybe not every individual agrees with this social contract, but we all experience the collective benefits (Di Mattia 2008).

Rawls's Contributions, a Path Paving Way to Social Ethics

Herein, it is particularly worth mentioning the ideas of John Rawls. In his *A Theory of Justice*, Rawls developed the theory of justice by describing the role of justice in social cooperation and by defining justice as fairness, and he dealt with the traditional conception of social contract (Rawls 1971, 1999). His elaboration of justice in the context of social cooperation and social contract paves the way to the idea of social ethics.

John Rawls (1921–2002), one of the most important political philosophers of the twentieth century, wrote highly influential articles in the 1950s and 1960s focusing on substantive problems of moral and political philosophy about what we ought to do. Rawls revitalized the social-contract tradition, using it to articulate and defend a detailed vision of egalitarian liberalism in his first book, *A Theory of Justice* (*TJ*) (1971). He recast the role of political philosophy, accommodating it to the effectively permanent "reasonable pluralism" of religious, philosophical, and other comprehensive doctrines or worldviews that characterize modern societies in Political Liberalism (1993). He explains how philosophers can characterize public justification and the legitimate, democratic use of collective coercive power while accepting that pluralism. *TJ* sets out and defends the principles of justice as fairness. Rawls takes the basic structure of society as his subject matter and utilitarianism as his principal opponent. Part 1 of *TJ* designs a social-contract-type thought experiment, the original position, and argues that parties in the original position will prefer justice as fairness to utilitarianism and various other views. In order to understand the argument from the original position, one must pay special attention to the motivation of the parties to the original position, which is philosophically stipulated and provided with a Kantian interpretation. Part 2 of *TJ* checks the fit between the principles of justice as fairness and our more concrete considered views about just institutions, thereby helping move us toward a reflective equilibrium that supports those principles. Part 3 of *TJ* addresses the stability of a society organized around justice as fairness, arguing that there will be an important congruence in such a society between people's views about justice and what they value (Richardson 2015).

In this context, Rawls developed the idea of public reason, as a conception of a well-ordered constitutional democratic society. To him, the form and content of this reason – the way it is understood by citizens and how it interprets their political relationship – is part of the idea of democracy itself. This is because a basic feature of democracy is the fact of reasonable pluralism; the fact that a plurality of conflicting reasonable comprehensive doctrines, religious, philosophical, and moral is the normal result of its culture of free institutions. Citizens realize that they cannot reach agreement or even approach mutual understanding on the basis of their irreconcilable comprehensive doctrines. In view of this, they need to consider what kinds of reasons they may reasonably give one another when fundamental political questions are at stake. Rawls argues that in public reason comprehensive doctrines of truth or right are replaced by an idea of the politically reasonable addressed to citizens as citizen. The well-ordered constitutional democratic society of Political Liberalism is one in which the dominant and controlling citizens affirm and act from irreconcilable yet reasonable comprehensive doctrines. These doctrines in turn support reasonable political conceptions although not necessarily the most reasonable which specify the basic rights, liberties, and opportunities of citizens in society's basic structure (Rawls 1997).

Rawls's conception of political liberalism provides valuable information about how a democratic society should deal with bioethical questions. The public use of reason consists, therefore, also of the development of procedures and criteria in order to apply the principles of justice and the political values they reflect (political and civic freedoms, equal opportunities, social equality, economic reciprocity, and general well-being) in a generally acceptable manner. The core of Rawls's political liberalism is this: we can reach an overlapping consensus on those political values that are part of a political notion of justice connected with a democratic basic order (Pauer-

Studer 2006). Rawls' contributions to moral philosophy have fertilized the new analyses on the ethical principle of justice and equality that is worth considering social ethics based on social contract and political liberalism.

Social Contract, Solidarity, Responsibility, and Social Ethics

The social contract tradition contends that society is established through a collective, mutually binding agreement or contract. Moral expectations and duties are shaped by the contract and its implications. Ethics is, therefore, primarily socially constructed and regulated. In a pre-political state, referred to as the state of nature, rational individuals accept to enter a reciprocal agreement because it is mutually beneficial. The contract is hypothetical in that its existence cannot be validated historically; however, its moral legitimacy derives from the assumption that rational, self-interested individuals would likely forge this agreement because they have more to gain from joining in a mutually beneficial association than from staying out of it. In the Hobbesian version of the social contract, morality, rooted in social reality, is a pragmatic, self-interested response to sustain survival. Other social contract theorists, such as Locke, Rousseau, and Rawls, do not adopt such a self-interested and pessimistic stance. Kantian contractarianism (or contractualism), for instance, values people as ends-in-themselves and assumes preexisting moral duties embedded in the human ability to reason (Rozuel 2013).

Conclusion

The idea of social ethics has been proposed to develop the themes of global governance, civil society, sustainable human development to point to the need of a partnership, and a long-term vision of our collective needs, a "sine qua non" condition for an equitable, hence, sustainable development process (Levy 1997). The ethical values and behaviors are not only abstract terms, but they are refined and conceptualized by the real-life experiences. The societal context where the actions of humans can be analyzed by ethical decision-making is entirely relevant to deliberate on what is the right thing to do, since the ethical values and principles response to the actual practices of life and to the needs of humans in the society. Furthermore the tension between individualism and communitarian needs can be reconciled with the perspective of social ethics by respecting the individual autonomy without disregarding the common good. Provided with basis of the ethical principles of non-maleficence, beneficence, autonomy, and justice, the idea of social ethics paves the way to a communitarian and social contractarian alternative to restore the injustices and inequalities in a democratic society fostered with the idea of public reason. By promoting the values of social responsibility, solidarity, and social utility, social ethics stands as the fulcrum of a rational, moral, egalitarian, pluralistic, democratic society rising on the pillars of human rights and human dignity.

Acknowledgment I would like to thank my colleagues, Nadi Bakirci, MD, PhD, professor of Public Health and Pinar Topsever, MD, PhD, associate professor of Family Medicine for their inspirations.

Cross-References

- ▶ Access to Healthcare
- ▶ Applied Ethics
- ▶ Benefit-sharing
- ▶ Bioethics: And Politics
- ▶ Biopolitics
- ▶ Citizenship
- ▶ Codes of Conduct
- ▶ Common Good
- ▶ Communitarian Bioethics
- ▶ Egalitarianism
- ▶ Equality and Equity
- ▶ Ethics
- ▶ Global Ethics
- ▶ Health Education and Promotion
- ▶ Health Insurance

- Health Policy
- Human Dignity
- Human Rights
- Moral Theories
- Poverty
- Professionalism
- Public Debate
- Public Health
- Right to Health
- Social Work
- Solidarity
- Welfare

References

Armstrong, A. C. (1907). Individual and social ethics. *The Journal of Philosophy, Psychology and Scientific Methods, 4*(5), 119–122. http://www.jstor.org/stable/2010586. Accessed 21 Sept 2014.

Beauchamp, T. L. (2003). Methods and principles in biomedical ethics. *Journal of Medical Ethics, 29*, 269–274.

Beauchamp, T. L. (2008). Ethical theory and bioethics. In T. L. Beauchamp, L. R. Walters, J. P. Kahn, & A. C. Mastroianni (Eds.), *Contemporary issues in bioethics* (7th ed., pp. 25–26). Belmont, USA: Wadsworth Cengage.

Beauchamp, T. L., & Childress, J. F. (2009). *Principles of biomedical ethics* (6th ed.). New York: Oxford University Press.

Childress, J. E., Faden, R. R., Gaare, R. D., Gostin, L. O., Kahn, J., Bonnie, R. J., Kass, N. E., Mastroianni, A. C., Moreno, J. D., & Nieburg, P. (2002). Public health ethics: Mapping the terrain. *Journal of Law, Medicine and Ethics, 30*, 170–178.

Craut, R. (2014). *Aristotle's ethics*. Stanford Encyclopedia of Philosophy. In E. N. Zalta (Principal Editor), First published Tue 1 May 2001; substantive revision Mon 21 Apr 2014, ISSN 1095–5054. http://plato.stanford.edu/entries/aristotle-ethics/. Accessed 26 Feb 2015

Di Mattia, P. (2008). Ethics. In W. Kirch (Ed.), *Encyclopedia of public health*. Springer Reference. www.springerreference.com. Berlin Heidelberg: Springer, 2011-01-31 23:00:00 UTC.

Gillon, R. (2003). Ethics needs principles – four can encompass the rest – and respect for autonomy should be "first among equals". *Journal of Medical Ethics, 29*, 307–312.

Hare, R. M. (2012). A utilitarian approach. In H. Kuhse & P. Singer (Eds.), *A companion to bioethics* (2nd ed., pp. 85–90). Chichester, West Sussex, UK: Wiley-Blackwell.

Lévy, B. (1997). Globalisation et développement: à la recherche d'une éthique sociale. *Canadian Journal of Development Studies/Revue canadienne d'études du développement, 18*(3), 519–534, Published online: 14 Jun 2013.

McCarthy, G. (1986). German social ethics and return to Greek Philosophy Aristotle and Marx. *Studies in Soviet Thought, 31*, 1–24.

Pauer-Studer, H. (2006). Public reason and bioethics. *MedGenMed, 8*(4), 13.

Rawls, J. (1971). *A theory of justice* (rev. ed.). Copyright 1999 by the President and Fellows of Harvard College, Printed in the United States of America. This book is a revised version of A Theory of Justice published in 1971 by Harvard University Press, Sixth Printing 2003.

Rawls, J. (1997). The idea of public reason revisited. *The University of Chicago Law Review, 64*(3), 765–807.

Rawls, J. (1999). *A theory of justice, Revised edition*. Cambridge, MA: Belknap Press of Harvard University Press.

+Richardson, H. S. (2015). John Rawls (1921–2002). *The Internet Encyclopedia of Philosophy* (A peer-reviewed academic resource). In J. Fieser, B. Dawson (Eds.), ISSN 2161–0002. http://www.iep.utm.edu/. Accessed 24 Feb 2015.

Rozuel, C. (2013). Social contract. In S. Idowu, N. Capaldi, L. Zu, & A. Gupta (Eds.), *Encyclopedia of corporate social responsibility*. Springer Reference. www.springerreference.com. Berlin\Heidelberg: Springer, 2013-03-13 15:49:25 UTC

Schwartz, L., Preece, P. E., & Hendry, R. A. (2002). *Medical ethics, a case-based approach*. Edinburgh, UK: Saunders Elsevier.

Shamoo, A. E., & Resnik, D. B. (2009). *Responsible conduct of research* (2nd ed.). New York: Oxford University Press.

Veatch, R. M. (2003). *The basics of bioethics* (2nd ed.). New Jersey: Prentice Hall.

Further Readings

Anand, S., Peter, F., & Sen, A. (Eds.). (2009). *Public health, ethics, and equity*. New York: Oxford University Press (Reprinted).

Buchanan, A. (2010). Health-care delivery and resource allocation. In R. M. Veatch (Ed.), *Medical ethics* (2nd ed.) (pp. 321–361). Mississagua, Canada: Jones and Bartlett Publishers.

Guttman, W. (2008). Ethics of health promotion. In K. Heggenhougen (editor-in-chief), & S. Quah (Associate editor-in-chief), *International encyclopedia of public health: Vol.* 2. (pp. 477–482). Academic Press is an imprint of Elsevier, Oxford, UK

Pogge, T. (2005). Human rights and global health: A research program. *Metaphilosophy, 36*(1–2), 2005.

Wikler, D., & Marchand, S. (2012). Society's allocation of resources for health. In H. Kuhse & P. Singer (Eds.), *A companion to bioethics* (Blackwell companions to philosophy 2nd ed., Vol. VIII, pp. 351–361). Chichester, West Sussex, UK: Wiley-Blackwell.

Social Work

Frederic G. Reamer
School of Social Work, Rhode Island College, Providence, RI, USA

Abstract

Social workers are critical participants in bioethics discussions, debates, and activities. Social workers serve as case managers, clinicians, agency managers and administrators, community organizers, policy analysts, advocates, and researchers. They address a wide range of ethical issues, such as end-of-life decisions, reproductive rights, management of confidential information shared by patients and family members, informed consent, minors' rights to confidentiality and treatment, patients' right to refuse treatment, paternalism, conscience clauses, and allocation of limited resources. This entry provides an overview of the social work profession, social workers' role in health-care settings, the evolution of ethical standards in social work, ethical challenges facing social workers in health-care settings, and relevant ethics concepts.

Keywords

Codes of ethics; Ethics; Professional ethics; Social work; Values

Introduction

Social workers throughout the world address issues of health care, poverty, mental illness, substance abuse, homelessness and affordable housing, eldercare, child welfare (protective services, adoption, foster care), disabilities, crime and delinquency, domestic and family conflict, trauma, and discrimination and oppression. Social workers serve as case managers, clinicians, agency managers and administrators, community organizers, policy analysts, advocates, and researchers. The roles social workers assume in various nations vary considerably. In some nations – particularly Western nations – significant numbers of social workers provide psychotherapeutic and other counseling services; in many nations, social workers serve primarily as case managers, advocates, and administrators.

The social work profession was inaugurated formally in the late nineteenth century, primarily in Europe and North America. Social work's earliest practitioners provided services in settlement houses, hospitals, psychiatric clinics, schools, and juvenile courts. Since then, the social work profession has broadened its reach to include a much wider range of settings and client populations.

Social Work in Health Care

Many social workers provide services in health-care settings, such as medical and psychiatric hospitals, outpatient medical facilities, hospice programs, rehabilitation facilities, physician group practices, skilled nursing facilities, nursing homes, home health agencies, and addiction facilities. Social workers in health-care settings provide services to patients, families, and groups to enhance their physical and emotional well-being. Social workers explain health-care resources and policies to patients, family members, and professional staff; help plan for post-hospital patient needs by arranging for services at another facility or in the home; explain to patients the causes of diseases, including environmental risk factors; help patients and families receive needed follow-up care by referral to health-care resources; understand and address social, cultural, and religious factors that contribute to patients' responses to illness and their use of health-care resources; provide advocacy through appropriate government and private-sector organizations; and help patients with advance directives and other long-term care issues.

Social workers have played a vital role in health-care settings since the early twentieth century. Social work was formally introduced to medical settings in the USA by Dr. Richard C. Cabot in 1905. Cabot, a professor of both clinical medicine and social ethics at Harvard University, was

instrumental in adding social workers to his clinic staff at Massachusetts General Hospital in Boston. Under the direction of their first department head, Ida Cannon, these social workers helped patients and their families cope with illness, disease, disability, and hospitalization by focusing particularly on their psychosocial needs, including their emotional reaction and adaptation (Kerson et al. 2010).

Over time, social work's function and influence in health-care settings have expanded significantly (Gehlert and Browne 2012), thus increasing the range of ethical issues they encounter. In addition to assisting hospitalized patients and their families, contemporary social workers provide genetic counseling, hospice services, psychotherapy and counseling in mental health agencies and outpatient health clinics that serve adults and children, and treatment of people with HIV-AIDS, eating disorders, and various addictions. These services sometimes involve complex ethical issues such as patient privacy and confidentiality, professional boundaries, informed consent, paternalism, truth-telling, and conflicts of interest. For example, a patient may disclose sensitive confidential information to a social worker concerning a mental health or addiction issue and ask the social worker not to share this information with family members or home health-care providers. Conversely, family members may share sensitive information with a social worker about the patient's complex personal history and ask the social worker not to share the information with the patient. Or, a terminally ill patient may want to talk to a social worker confidentially about her suicidal thoughts and end-of-life preferences. Such situations require complex management of sensitive ethical issues.

Social workers are also skilled in organizing and facilitating support groups for various populations, such as cancer patients, molestation and sexual assault victims, and parents of seriously impaired infants. They work to enhance the availability of community-based resources (e.g., support groups, health-care clinics in low-income neighborhoods, residential services), advocate on behalf of individual patients who are in need of services, and advocate to ensure that important public policy issues related to health care are addressed (e.g., health-care subsidies for low-income patients and family members who provide them with home-based care, coverage for mental health and addiction services). These roles also lead to challenging ethical issues. For example, social workers who facilitate support groups must be familiar with unique ethical standards pertaining to the confidentiality rights and obligations of group participants. Social workers who participate in the development of guidelines concerning involuntary psychiatric commitment must be familiar with ethical standards concerning informed consent and patients' right to refuse treatment. Social workers who seek to establish residential addiction facilities or HIV-AIDS programs may find themselves caught between the competing rights of clients and neighborhood residents who oppose locating such programs in their community.

In addition to ethical issues social workers encounter when they provide clinical and advocacy services, health-care social workers also face ethical challenges in their roles as program administrators, policy professionals, and members of institutional ethics committees and institutional review boards. Social work education's broad focus on clinical, administrative, advocacy, and policy knowledge and skills is particularly useful when practitioners encounter ethical issues in health-care settings.

Social workers' efforts to develop conceptual frameworks to address ethical issues began especially in the 1970s, as the broader field of applied, professional, and practical ethics developed. These efforts were particularly prominent with respect to social work in health-care settings, due largely to the emergence of bioethics as a specialized area of knowledge and professional practice. However, ever since social work's emergence in the late nineteenth century, practitioners have focused on a variety of ethical and moral issues.

The Evolution of Social Work Ethics

Social workers' exploration of ethical issues, including those related to health care, has evolved

over time and includes several conceptually distinct, albeit sometimes overlapping, periods. These include the morality period, values period, ethical dilemmas and decision-making period, risk management period, and digital period.

Social work's historical literature suggests that for many years, especially during the late nineteenth and early twentieth centuries, the profession was focused much more on the morality of clients (often known as paupers) than practitioners. In many scholarly discussions, one finds references to concern about the moral fiber, or the alleged lack thereof, of clients who struggled with issues such as poverty, unemployment, alcohol use, mental illness, or poor health. The phrase "professional ethics" did not exist during this period. Some of the discussions of clients' morality had a rather paternalistic tone.

As social work matured as a profession, a handful of scholars and practitioners began exploring and writing rich commentaries about the profession's core values, such as client dignity, self-worth, self-determination, and confidentiality. These important discussions and analyses sought to explore the implications of social work's central values, especially when there were conflicts among the profession's values, the broader society's values, and social workers' personal values. Especially during the turbulent 1960s and early 1970s, several social work authors wrote about the complex connections between social work values and contemporary controversies surrounding civil rights, women's rights, welfare rights, prisoners' rights, discrimination, and abortion (Banks 2012; Dolgoff et al. 2012; Hugman 2013; Reamer 2012). These discussions and debates were especially prominent in health-care settings with respect to clients' right to refuse treatment, informed consent related to medical treatment, and women's reproductive rights.

In the late 1970s, a relatively small group of social work scholars and practitioners began to pay close attention to the nature of ethical dilemmas in the profession. This development was an outgrowth of the dramatic emergence and maturation of the broader field of applied and professional ethics, especially bioethics. During this period, increasing numbers of scholars and practitioners in a wide range of professions (e.g., medicine, nursing, psychology, social work, journalism, business, law enforcement, engineering) focused explicitly for the first time in their respective histories on the nature of challenging ethical dilemmas facing practitioners. Discussions frequently focused on conflicts among professionals' duties and obligations. Prominent examples in social work included conflicts between clients' confidentiality rights and social workers' duty to disclose confidential information without client consent to protect third parties; management of complex dual relationships and conflicts of interest; and imposing limits on clients' right to self-determination and autonomy. In health-care settings, social workers focused especially on ethical dilemmas related to end-of-life decisions, reproductive rights, management of confidential information shared by patients and family members, informed consent, minors' rights to confidentiality and treatment, patients' right to refuse treatment, paternalism, conscience clauses, and allocation of limited resources (such as health-care funds and scarce organs).

During this period, scholarship burgeoned on the subjects of ethical dilemmas in practice and ethical decision-making protocols. The richest discussions identified links between ethical theory, drawn from the discipline of moral philosophy, and real-life challenges faced by professionals, particularly those involving conflicts among professional duties and obligations.

For the first time in social work's history, textbooks included in-depth overviews of ethical dilemmas and conceptual frameworks practitioners could use to address them. In social work, common topics concerned professional paternalism, the limits of clients' confidentiality rights, managing informed consent challenges, complicated boundary issues, dual relationships, conflicts of interest, allocation of limited resources, whistle-blowing, and compliance with allegedly unjust laws. Over time, social work education programs developed curricula to teach students about ethical dilemmas and decision making.

In the early 1990s, yet another trend emerged, especially in the USA, which, compared with

many other nations, is generally known as a relatively litigious society. This trend continues today. Although many social workers have sustained their interest in ethical dilemmas and decision making, new concerns emerged regarding ethics-related risk management. Data began to circulate concerning increases in lawsuits and licensing board complaints that raise ethical issues. Until this period, social workers rarely discussed such risks. Increased publicity have alerted social workers to relatively new information about the ways in which their ethical judgments could lead to litigation and licensing board complaints. Social workers have discovered how disgruntled clients and others could file formal complaints alleging, for example, mismanagement of clients' confidential and privileged information, boundary and dual relationship violations, conflicts of interest, negligent service delivery, fraud, and inappropriate termination of services.

For the first time in social work's history, literature has emerged about the links between social workers' ethical judgment and potential malpractice, negligence, and professional discipline (Reamer 2014). Licensing boards in some nations, especially the USA and Canada, have developed websites listing social workers who were sanctioned because, for example, they committed fraud, had sex with clients, and disclosed sensitive confidential information without proper authorization.

Professional ethics no longer is limited to questions such as, "What's the right thing to do in this complicated situation?" For many social workers, ethics now includes questions such as "Can I be sued or have a licensing board complaint filed against me if I. . .?" Ethics-related risk management has become a relatively new component of social work education and training. In health-care settings, social workers have been concerned about making ethical judgments in ways that protect clients and prevent lawsuits and ethics.

Today's social workers in health-care settings can provide clinical services online or via video counseling or telephone to clients they never meet in person. They may receive Facebook friend requests from clients or former clients that lead to boundary challenges. In this digital era, social workers can use e-mail and text messages as therapeutic tools and provide clients with specialized smartphones that enable them to record and transmit summaries of their deeply personal health symptoms and moods to their clinicians and case-workers, receive therapeutic messages and alerts, and communicate with other people in their digital network who face similar health challenges (a virtual support group).

The advent of this technology in health care has introduced unprecedented and controversial ethical challenges for social workers related to informed consent, privacy and confidentiality, boundaries, documentation, client abandonment, and the delivery of services across jurisdictional borders. Licensing boards and regulatory bodies in a number of nations are actively engaged in earnest efforts to develop constructive guidelines to govern entirely new forms of social work practice.

The Social Work Role in Bioethics

Social workers offer a unique perspective on bioethical challenges in health-care settings because of the profession's broad-based and explicit focus on clinical, administrative, policy, social justice, cultural diversity, and ethical issues (Banks 2012; Dolgoff et al. 2012; Hugman 2013; Reamer 2012, 2013). As clinicians, social workers can help clients and family members cope with difficult moral judgments related, for example, to end-of-life care, genetic screening, pregnancy management, consent to treatment, and termination of services. As administrators, social workers can participate in difficult ethical judgments about the allocation of limited and scarce health-care resources (known in moral philosophy as issues of distributive justice), implementation of controversial health-care laws and regulations, management of ethics crises and unethical practices, and the implementation of patients' rights policies. As community organizers and advocates, social workers can help stakeholders navigate orchestrated challenges to controversial health-care policies and laws and manage complex conflicts concerning health-care financing. As policy

professionals, social workers can contribute to discussions and debates about the ethical implications of health-care reforms, especially regarding access to health care for vulnerable and oppressed populations. That is, social workers are trained to examine ethical issues through multiple lenses, focusing simultaneously on pertinent clinical, policy, administrative, social justice, and cultural diversity issues.

Social workers in health-care settings typically function as part of an interdisciplinary team, which may include physicians, nurses, occupational and physical therapists, nutritionists, rehabilitation staff, clergy, and health-care administrators. On occasion, social workers facilitate the process through which health-care professionals negotiate differences of opinion or conflict among themselves concerning specific ethical issues. In this respect, social work's unique mission, perspectives, and methods can be especially useful when health-care professionals encounter ethical challenges. Social workers are trained to mediate conflict and understand that clinical ethics (e.g., decisions about access to services, end-of-life care, reproductive rights, organ donation, and patient privacy) is linked inextricably with challenging policy and social justice issues. Hence, the social work perspective – which calls for simultaneous examination of relevant clinical, administrative, policy, cultural diversity, religious, social justice, and ethical issues – is particularly valuable in interdisciplinary health-care settings where colleagues (especially physicians, nurses, and other allied health professionals) may be inclined to focus primarily on clinical issues.

Using their clinical skills, social workers can be particularly helpful when there is complicated intersection between ethical issues and patients' mental health challenges and family dynamics. For example, social workers' skilled use of mediation techniques can help to resolve disagreements that sometimes arise in health-care settings between patients and family members. Often, what appear to be intense ethics-related disagreements among family members – for example, decisions about honoring a patient's wishes regarding end-of-life care or telling a relative the truth about a grim prognosis – may reflect long-standing family conflicts; skilled clinical social workers who are sensitive to ethical issues can help family members and health-care professionals resolve complex ethical challenges by identifying and addressing relevant family issues. Social workers use their unique training and clinical skills to help patients, family members, and colleagues understand how some ethical issues intersect with mental health issues and family dynamics. Further, social workers understand the ways in which patients' mental health challenges (e.g., chronic and debilitating depression or anxiety) may influence the patients' management of ethical choices pertaining to end-of-life care, disclosure of confidential information, refusal of treatment, and informed consent.

In the policy arena, social workers are critically important participants in discussions of health-care reform. Debate among politicians, health-care advocates, citizens, insurance industry executives, and health-care administrators about health-care benefits and coverage, eligibility criteria, and access to services ultimately has moral implications. Social workers' participation in these conversations can increase the likelihood that participants will appreciate and address the ethical implications of their policy decisions for patients, families, communities, and the broader society (Holtz 2008).

Social work's traditional focus on cultural and religious diversity is especially relevant when ethical issues arise in health-care settings around the globe. Social workers' sensitivity to ethnic, cultural, and religious issues can be particularly helpful when there is a clash between the patients' and families' values and prevailing ethical norms, health-care policies, and practices (e.g., concerning the acceptability of certain lifesaving interventions, organ donation and transplantation, mood-altering medication, autopsy, or blood transfusion). Social workers are trained to understand that key bioethics concepts such as informed consent, privacy, confidentiality, boundaries, paternalism, truth-telling, and autonomy are interpreted very differently in diverse international communities and across ethnic, cultural, social, and religious groups. For example, some

cultures value autonomy and patient self-determination, whereas others are more likely to defer to health-care professionals' authority and judgment. Some cultures insist on strict boundaries between the practitioners' and clients' lives, while others are much more flexible and appreciate well-meaning dual relationships. In the USA and Canada, for example, both laws and ethical norms emphasize the clients' right to control disclosure of confidential information. In contrast, in some Asian nations, ethical norms permit and encourage social workers to share sensitive information with the clients' family members even though clients have not provided formal consent for the release of this information (Marsiglia and Kulis 2009; Reamer 2012).

Some of the bioethical issues that social workers encounter require specialized knowledge pertaining to specific medical conditions (Kerson et al. 2010). Examples include ethical dilemmas related to a family's decision about withdrawal of a cancer patient's life support, abortion following a rape, organ transplantation, the use of restraints with a noncompliant psychiatric patient, or a patient's decision to refuse neuroleptic medication. When such issues arise, social workers often serve as important intermediaries in relationships among patients, their families, and health-care professionals. In these instances, social workers help patients and their families make difficult personal decisions, facilitate communication among members of the health-care team, advocate on a patient's or family's behalf, or raise policy issues that need to be addressed by a hospital, nursing home, home health agency, or rehabilitation facility.

Other bioethical issues concern the nature of relationships and transactions between social workers and patients or their families. For example, social workers in health-care settings must be familiar with privacy, confidentiality, informed consent, and boundary-related norms that govern relationships with patients and families. They must also be sensitive to complex ethical issues involving patients' right to self-determination, truth-telling, professional paternalism, and whistle-blowing).

In particular, social workers can clarify differences among the different, sometimes conflicting, ethical obligations that guide various health-scare professions. For example, social workers in a health-care setting can help clarify the ethical responsibilities of various professionals when minors request health-care services without notification of their parents or when staff suspect child abuse or that a patient with AIDS poses a threat to a sexual partner. Laws in different jurisdictions vary, for instance, regarding minors' right to health care and regarding which health-care professionals are permitted to disclose confidential information, without patients' consent, to protect third parties who are at risk of exposure to HIV or AIDS (Slater and Finck 2012). These laws differ among nations and among jurisdictions within nations.

In addition to their clinical role, social workers may be members of institutional ethics committees (IECs) that discuss ethically complex cases and health-care policies (Hester and Schonfeld 2012; Post et al. 2007). They may offer a particularly valuable perspective because of their extensive contact with patients and their families and can, therefore, contribute to discussions about, for example, resuscitation guidelines, patients' right to refuse treatment, advance directives, organ transplantation, treatment of severely impaired infants, patients' privacy rights, and patients' participation in high-risk research protocols. Because of their broad-based education, social workers who serve on ethics committees can offer both clinical and policy-related expertise, particularly when issues arise concerning the implications of cultural diversity (including ethnicity, race, disability, gender identity and expression, sexual orientation, and religion) for ethical judgments and policies.

The concept of ethics committees (also known as institutional ethics committees) first emerged in 1976, when the New Jersey Supreme Court in the USA ruled that Karen Ann Quinlan's family and physicians should consult an ethics committee to help them decide whether to remove Quinlan from life-support technology. Ethics committees, which have been most prominent in health-care settings (especially hospitals, nursing homes, rehabilitation facilities, hospice, and home health-care programs), typically include

representatives from various disciplines and positions, such as nursing, medicine, social work, the clergy, and agency administration. (There is some debate about whether an agency's attorney should be on an ethics committee because of lawyers' unique fiduciary duty to protect their clients' interests first and foremost.)

Some ethics committees include an ethicist – either an agency employee (for instance, in large teaching hospitals) or an outside consultant – who has formal training in applied and professional ethics, moral philosophy, and ethics consultation. Some ethicists are trained philosophers or theologians with a special interest in professional ethics and bioethics, and some are members of a human services profession (such as nursing, social work, or medicine) who have supplemental education related to ethics.

Many ethics committees provide agency staff with case-related consultation services and nonbinding advice, particularly when staff members or clients want assistance in thinking through difficult ethical decisions. For example, in hospital settings, ethics committees may offer consultation and nonbinding advice on issues related to termination of life-support technology, the use of aggressive care with terminally ill patients, patients' right to refuse treatment, and patients' eligibility for organ transplantation. Although ethics committees are not always able to provide definitive advice or guidance about complex ethical issues, they can offer colleagues and clients with a forum for organized, focused, explicit, and principled exploration of ethical dilemmas. This can provide participants with a greater understanding of the issues and options they face and enhance the quality of their decision making.

Many ethics committees also serve other functions. Some are responsible for reviewing existing ethics-related policies in health-care settings and suggesting revisions, sometimes in response to controversial case-related issues that arise in the agency. Ethics committees also draft new ethics-related policies and procedures for more formal review and approval by agency administrators and boards of directors.

Some social workers are active participants on institutional review boards (IRBs) in health-care settings that examine a variety of ethical issues in research involving human participants. Ethical standards in social work pertaining to research include extensive guidelines concerning evidence-based practice, informed consent, privacy and confidentiality, protection from harm, and conflicts of interest.

In addition, social workers may be involved in workplace discussions and legislative advocacy about the ethical aspects of health-care financing mechanisms and cost-containment measures. They may also propose ways to advocate on patients' behalf or to advocate for policy reform in the public and private sectors that may provide a more just allocation of scarce health-care resources at the local, national, or international level. An example is social workers' participation on a task force whose mission is to enhance low-income people's access to basic health care or on a hospital committee concerned about pressure to limit care provided to, and hasten discharge of, psychiatric patients covered under managed care programs operated by private, for-profit insurers. In these instances, social workers may help identify the psychosocial consequences of various strategies to allocate limited health-care resources.

Enhancing Social Work's Contribution to Bioethics

To participate fully in discussions of bioethical issues and dilemmas, social workers need specialized knowledge and training. First, they need to be familiar with the history, language, concepts, and theories of bioethics as they pertain to practical ethics challenges, particularly as they have evolved since the early 1970s, when the professional ethics field began in earnest (Copp 2007). Second, social workers should be knowledgeable about formal mechanisms that can help health-care professionals monitor and address bioethical issues. These include phenomena such as institutional ethics committees, institutional review boards, utilization review and quality assurance committees, informed consent procedures, and advance directives. Social workers should understand the nature of ethics consultation in health-care settings and the ways in which ethicists can

serve as useful resources. In addition, social workers should participate in policy-making efforts in health-care settings concerning such issues as conscience clauses, advance directives, organ donation and transplantation, patients' rights, do-not-resuscitate orders, withholding or withdrawing life-sustaining treatment, refusal of blood transfusion, stem cell research, and research involving human participants.

It is also useful for social workers to be acquainted with relevant codes of ethics and legal considerations (statutes, regulations, case law) related to patients' rights and health-care professionals' obligations. In many nations, codes of ethics in social work and allied health professions have become increasingly sophisticated with regard to issues such as informed consent, confidentiality, privacy, privileged communication, conflicts of interest, dual relationships, client abandonment, allocation of limited resources, compliance with unethical policies and orders, dishonesty and fraud, and impaired professionals (Reamer 2006).

The breadth and depth of social work codes of ethics around the world vary considerably. Social work codes of ethics exist in three different contexts. First, internationally, many professional social work associations have developed codes of ethics. These are voluntary associations of social workers. Thus, these codes typically carry no legal or formal regulatory authority, although the associations may insist on compliance with the code as a condition of membership. The International Federation of Social Workers publicizes ethics codes from Australia, Canada, Denmark, Finland, France, Germany, Ireland, Israel, Italy, Japan, Luxembourg, Norway, Portugal, Russia, Singapore, Spain, Sweden, Switzerland, Turkey, the USA, and the UK.

Secondly, in some nations, social work ethics codes have been developed or adopted by governmental licensing boards or regulatory bodies that authorize social work practice. Some governmental licensing boards and regulatory bodies develop their own ethical standards, some formally adopt portions of codes developed by prominent voluntary social work associations, and some formally adopt entire codes developed by prominent voluntary social work associations. These ethical standards become legally enforceable once they are incorporated into licensing statutes or regulations.

Thirdly, many private-sector human and social service agencies have adopted codes of ethics or ethical standards to which employees are bound. Here too, some of these organizations develop their own unique ethical standards and some draw on codes of ethics developed by prominent voluntary professional associations. In some instances – especially when the agency hires employees educated in different professions – agencies will draw on ethical standards from diverse professions, choosing those standards that are most relevant to their mission, client population, programs, services, and local cultural norms.

Social work codes of ethics are remarkably diverse in their purpose, content, and format. Several prominent examples illustrate this diversity. For example, in the UK, the British Association of Social Workers (BASW), a voluntary association, offers a code of relatively modest length that defines social work and provides an overview of core values and principles and guidelines for ethical practice focusing on "service users," the social work profession, the workplace, and social workers' responsibilities in particular roles (management, education, training, supervision and evaluation, independent practice, and research). Since membership in the BASW is voluntary, this code is not used for regulatory purposes.

In England, a mandatory code of conduct for social workers has been adopted by the General Social Care Council (the body appointed by the UK government to regulate professional social work and professional social work education). As a result, the term "social worker" became a protected title. Social workers are required to accept a code of conduct and could be disciplined if their actions are deemed to have breached the code. Prior to implementation of this code of conduct, any person in the caring professions could use the term social worker irrespective of role or qualification. The English code includes three broad sections: an introduction that

describes the function of the guidelines, a code of practice for employers, and a code of practice for social care workers.

Similar to the BASW, the Canadian Association of Social Workers (CASW) offers a widely disseminated code of ethics to which members are held. In contrast to the English approach, which enforces ethical standards through a national code of practice that applies to a diverse group of "social care workers," the individual Canadian provinces have the authority to develop and adopt legally enforceable codes of ethics that pertain exclusively to social workers. The CASW Code of Ethics is a relatively short document; it includes a brief preamble and a summary of core social work values and principles related to respect for inherent dignity and worth of persons, pursuit of social justice, service to humanity, professional integrity, confidentiality, and competence. The CASW also publishes a companion document, *Guidelines for Ethical Practice*, which offers detailed guidelines regarding a wide range of ethical issues and challenges.

In Canada, enforceable ethical standards exist at the provincial level. Thus, the Alberta College of Social Workers, the Ontario College of Social Workers and Social Service Workers, and the Newfoundland and Labrador Association of Social Workers, among other regional groups, have the authority to develop their own standards of practice.

In yet another arrangement, the principal set of ethical standards in Australia is promulgated by the Australian Association of Social Workers (AASW), a voluntary professional association. In Australia, the AASW code is the national lodestar, even though the organization does not have the legal authority to regulate social work. Some social workers choose not to join AASW and, therefore, fall outside the AASW's purview. The AASW code includes a summary of core values, ethical standards, and a brief guide to ethical decision making. The AASW also publishes extensive practice standards with which members are expected to adhere in conjunction with the code of ethics.

In contrast, the Singapore Association of Social Workers Code of Ethics includes a mix of general principles and narrowly focused ethical standards concerning social workers' ethical responsibilities to clients and colleagues, in practice settings, to the social work profession, and to the broader society. This is an example of a code adopted by a voluntary professional association that incorporates abstract principles concerning social work's mission and narrowly worded standards concerning such topics as audio recording of clients, sexual misconduct, and informed consent by minors.

The National Association of Social Workers (USA) Code of Ethics has served as an influential model and resource for many codes of ethics around the world. This detailed code includes a preamble and mission statement for the social work profession, a summary of core values and broad ethical principles, and a comprehensive list of specific ethical standards. A number of social work associations, licensing boards, regulatory bodies, and private social and human service agencies have drawn on the NASW code. Within the USA, the NASW Code of Ethics is the best known and most influential set of ethical standards applying to social workers. Ethics codes have also been developed by other social work organizations in the USA, such as the National Association of Black Social Workers and the Clinical Social Work Association.

In the USA, codes of ethics are used by several bodies that govern social workers. The NASW, the nation's largest professional social work association, uses the NASW Code of Ethics to review and adjudicate ethics complaints filed against NASW members. Further, many of the individual state licensing boards in the USA have formally adopted the NASW Code of Ethics, or portions of the code, and use it to review and adjudicate ethics complaints filed against licensed social workers. Finally, many public and private social service agencies have adopted the NASW Code of Ethics or portions of the code, as formal agency policy, and use the code's standards to guide and assess employees' conduct.

In addition to the codes of ethics promulgated by voluntary social work associations, licensing and regulatory bodies, and private agencies in many nations, the code developed by the

International Federation of Social Workers (IFSW) and International Association of Schools of Social Work (IASSW), *Ethics in Social Work: Statement of Principles*, provides an overarching, truly international code of ethics. The IFSW represents social work organizations in some 90 nations and places special emphasis on human rights, human dignity, social justice, and professional conduct. The IASSW includes members from educational institutions throughout the world. This code was written deliberately at a fairly high level of abstraction. The IFSW-IASSW code makes explicit reference to relevant international human rights declarations and conventions, including the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights.

It is essential for social workers to appreciate the ways in which their involvement in, and view of, ethical issues vary in different nations and cultures. Concepts such as informed consent, privacy, confidentiality, and professional-client boundaries have very different meanings in different cultural contexts. Issues pertaining to the allocation of health-care resources are profoundly different in countries with nationalized health care and countries where health-care services are provided by for-profit and nonprofit private-sector hospitals and outpatient clinics. In nationalized health-care systems, resource allocation decisions are not based on implications for investors or shareholders; however, decisions in for-profit health-care settings are likely to be guided by such considerations.

Finally, social workers should be familiar with the various schools of thought and bioethics concepts that pertain to ethical decision making and ethical theory (including relevant theories of normative ethics, such as deontology, utilitarianism, virtue ethics, and care ethics). This can be particularly useful when social workers are involved in discussions of cases with professional ethicists, for example, when a decision must be made about when and how to tell a fragile, terminally ill patient the truth about his or her diagnosis or to disclose confidential information, against a patient's wishes, in order to protect a third party. This training may be offered as part of agency-based in-service education, professional conferences, or undergraduate and graduate social work education.

Conclusion

Especially since the 1970s, social workers throughout the world have been aware of the diverse and complex bioethical issues involved in health care, whether they involve acute or chronic, inpatient or outpatient, or medical, rehabilitative, nursing, or psychiatric care. Social workers' growing awareness of, and enhanced expertise in addressing, global bioethical issues helps to ensure the protection of patients' and families' rights and the soundness of ethical decisions made in health-care settings.

Cross-References

- ▶ Applied Ethics
- ▶ Care Ethics
- ▶ Clinical Ethics: Consultation
- ▶ Codes of Conduct
- ▶ Committees: Clinical Ethics Committees
- ▶ Confidentiality
- ▶ Consultation
- ▶ Ethics
- ▶ Malpractice
- ▶ Mental Health: Services
- ▶ Moral Theories
- ▶ Professional Ethics
- ▶ Research: Human Subjects
- ▶ Standards of Care
- ▶ Utilitarianism

References

Banks, S. (2012). *Ethics and values in social work* (4th ed.). Hampshire: Palgrave Macmillan.

Copp, D. (Ed.). (2007). *The Oxford handbook of ethical theory*. New York: Oxford University Press.

Dolgoff, R., Harrington, D., & Loewenberg, F. (2012). *Ethical decisions for social work practice* (9th ed.). Belmont: Brooks/Cole.

Gehlert, S., & Browne, T. (Eds.). (2012). *Handbook of health social work* (2nd ed.). Hoboken: Wiley.

Hester, M., & Schonfeld, T. (Eds.). (2012). *Guidance for health care ethics committees*. Cambridge: Cambridge University Press.
Holtz, C. (2008). *Global health care: Issues and policies*. Sudbury: Jones and Bartlett.
Hugman, R. (2013). *Culture, values, and ethics in social work: Embracing diversity.* Abingdon, Oxon: Routledge.
Kerson, T., McCoyd, J., & Associates. (2010). *Social work in health settings: Practice in context* (3rd ed.). New York: Routledge.
Marsiglia, F., & Kulis, S. (2009). *Diversity, oppression, and change: Culturally grounded social work*. Chicago: Lyceum.
Post, L., Blustein, J., & Dubler, N. (2007). *Handbook for health care ethics committees*. Baltimore: Johns Hopkins University Press.
Reamer, F. (2006). *Ethical standards in social work: A review of the National Association of Social Workers code of ethics* (2nd ed.). Washington, DC: NASW Press.
Reamer, F. (2012). *Boundary issues and dual relationships in the human services*. New York: Columbia University Press.
Reamer, F. (2013). *Social work values and ethics* (4th ed.). New York: Columbia University Press.
Reamer, F. (2014). *Risk management in social work: Preventing professional malpractice, liability, and disciplinary action*. New York: Columbia University Press.
Slater, L., & Finck, K. (2012). *Social work practice and the law.* New York: Springer.

Further Readings

Corey, G., Corey, M., Corey, C., & Callanan, P. (2015). *Issues and ethics in the helping professions* (9th ed.). Stamford: Cengage.
Dziegielewski, S. (2013). *The changing face of health care social work* (3rd ed.). New York: Springer.
Kirst-Ashman, K. (2013). *Introduction to social work and social welfare* (4th ed.). Belmont: Brooks/Cole.
Zastrow, C. (2014). *Introduction to social work and social welfare* (11th ed.). Belmont: Brooks/Cole.

Solidarity

Sally Scholz
Philosophy Department, Villanova University, Villanova, PA, USA

Abstract

This entry first presents some of the definitions and theoretical approaches to solidarity. Then, five different uses of solidarity within global bioethics are discussed: public health, welfare, universal health care, a right to health care, and a focus on the vulnerable.

Keywords

Collective responsibility; Right to health care; Solidarity; Universal health care; Welfare state

Introduction

The concept "solidarity" may be usefully contrasted with both "charity" and "justice," although the three are also intimately related both historically and conceptually. Whereas justice generally pertains to just distributive or retributive arrangements such as "fairness" or "maximization of utility" (among other things), "charity" connotes the obligation of privileged individuals to aid those in need, whether the need be material, physical, or emotional. "Solidarity" involves elements of both in referring to the willingness to acknowledge social bonds with others and to act on behalf of the needs or interests of those others, whether or not those needs and interests are shared.

Solidarity in moral and political philosophy is generally understood to fall within the literature on collective responsibility and may be understood descriptively or normatively. As a descriptive concept, the term denotes the extent of cohesion in a community or group. As a normative concept, solidarity connotes a number of possible meanings from a principle that motivates justice-seeking behavior to a relation that informs particular duties. After a brief presentation of the definition of the concept in different contexts, solidarity as a moral relation and solidarity as a moral principle are discussed. Then, five different uses of solidarity within global bioethics are presented, highlighting accounts that appeal to solidarity as a moral relation or a moral principle: public health, welfare, universal health care, a right to health care, and a focus on the vulnerable.

Defining Solidarity in Context

A wide variety of definitions have been offered for solidarity, and an expanding array of theoretical approaches explains its role in moral and political theory. The very general definition above attempts to capture what is found in most definitions: a form of unity or social bond and a requisite action in response to that unity. Most definitions also posit an element of personal transformation, that is, individuals in solidarity are transformed by their participation in the solidaristic group or activity. The theoretical approaches add nuance to the earlier classification of descriptive and normative frameworks. Three basic types emerge: social solidarity, political solidarity, and civic solidarity.

Social solidarity was made famous by Emile Durkheim who described the social relations of different forms of society. A society marked by relative similarities between members is said to exhibit mechanical solidarity. In contrast, more advanced societies exhibit what Durkheim called organic solidarity which is characterized by an extensive division of labor. Durkheim described morality emerging out of these relations of solidarity. Social solidarity, then, is the social cohesion among participants. Political solidarity is the solidarity associated with social movements against injustice. Unlike social solidarity, political solidarity does not rely on or emphasize the bonds between group members but rather the actions they take in concert with one another to effect change. Finally, civic solidarity is the solidarity between citizens and their formal organized governments; the term here is used to denote the state's obligations to protect citizens against vulnerabilities, especially those vulnerabilities that arise from or are intensified by social existence. These three theoretical approaches to solidarity yield many different manifestations of solidarity thereby lending to the wide and diffuse usage of the term in social and political practice (see Bayertz 1999; Scholz 2008).

One way to highlight the distinctions among the various definitions and theoretical approaches to solidarity is to note how they address obligations of solidarity. Social solidarity, which pertains to the cohesion of a community, appears to establish a particularistic framework wherein the obligations are determined by the nature, extent, and aims of the particular community. Families have obligations to members that are particular to that social unit; we would not expect strangers to have the same sorts of obligations that we expect of parents or siblings (and indeed even those relations inform particular duties). Sports teams, classmates, members of professional societies, and countless other groupings or communities have relations of social solidarity informed by the specific nature, extent, and aim of the community. Social solidarity might also extend to all humanity, sometimes called human solidarity, in which case solidarity is not particularistic but universal. In this latter case, the obligations extend to all humanity but pertain to the bonds that tie human beings together. In other words, there might be certain obligations of solidarity that all human beings have that are not readily conceived under rubrics of justice or charity. An example of such a solidaristic obligation is the obligation to recognize another human being as part of the collective humanity.

Civic solidarity is particular to a given society; the entire community is obliged to protect each citizen, and each citizen in turn must also ensure the community's ability to work for the common good. With cooperative international organizations, it is possible to conceive of a civic solidarity functioning transnationally. For instance, the coordinate multinational efforts to stem the outbreak of SARS (severe acute respiratory syndrome) or MERS (Middle East respiratory syndrome) may be held up as examples of global civic solidarity insofar as regional and international organizations were responding to obligations to protect vulnerable populations against some of the health hazards accompanying globalized relations.

Obligations of political solidarity are shaped by those actions that advance the overall social justice aim of the group. However, that aim in and of itself informs actions toward one another and toward those outside the group (often those individuals or societies the solidaristic group is trying to change).

As may be evident, solidarity in bioethics is primarily a form of civic solidarity. A further distinction helps in analyzing the varying uses of solidarity in bioethics. This distinction is between solidarity as a moral relation and solidarity as a principle (a third distinction, solidarity as a disposition, may also be identified but it is arguably an affective facet of the principle within bioethics literature (see, for instance, Houtepen and ter Meulen 2000b)). In the first instance, the concept refers to the solidaristic group or the relationship between solidary actors. In the second, solidarity is a truth or claim that determines or informs a series of moral actions, often these actions are undertaken by individuals, but at times the principle of solidarity clearly refers to state-based action.

Solidarity as a Moral Relation

Accounts of the moral relation of solidarity vary according to the type or basis of the bonds that bring people together into a solidary community. Three common accounts of those bonds include commonality of physical attributes, shared experiences, and mutual interests.

Commonality of attributes is used to assert or imply connections between individuals who outwardly appear to resemble each other. For instance, peoples of similar gender or race may presume a connection with others of that gender or race simply based on the commonality. Similarly, shared experiences, like surviving cancer or being at risk for serious disease, can inspire relations of solidarity wherein the individual members not only sense a connection to one another but act on that connection. Finally, mutual interest can unite otherwise dissimilar people. This is often considered the basis of welfare state politics and hence plays an important role in conceptions of solidarity for bioethics. In each of these, it might be argued that those in solidarity exhibit partiality toward similarly situated others. That may indeed be the very meaning of the solidarity, i.e., a particular relation informed by commonality that allows partiality in moral decision-making. However, partiality may not be the only or even an accurate assessment of the solidaristic relation. Solidarity as a moral relation may also be understood as inclusive, taking into account the interconnections within community into account while also valuing diversity or difference among individuals (Baylis et al. 2008).

These accounts of the relation of solidarity capture elements frequently associated with everyday language use of the concept. Mutuality, reciprocity, recognition, and allegiance to others fill in at least some of the content of the concept of solidarity. In varying ways, they demonstrate the place of solidarity within relational ethics more generally.

Solidarity as a Principle

As a principle, solidarity is set as a rule or truth which informs other obligations, including commitment to the common good and personal responsibility, or perhaps even more extensive moral and political systems. The principle of solidarity asserts the connectedness or interdependence of human beings as an empirical fact and seeks to achieve other such fundamental values as justice and equality.

As a principle in bioethics, solidarity is sometimes articulated as a counter to justice and autonomy (Hoedemaekers and Dekkers 2003; Butler 2012). Whereas autonomy emphasizes the individual, solidarity emphasizes community or the connections between members of communities. It is this sense that appears prominently in the European tradition generally and European bioethics and biolaw specifically. In the United States, the principles of bioethics are autonomy, nonmaleficence, beneficence, and justice. Solidarity sometimes explicitly replaces the emphasis on autonomy or augments dignity and vulnerability in a revised list of the basic principles: autonomy, dignity, integrity, and vulnerability (Sass 2001, p. 219). Stone (2012) argues that principles of solidarity and care, which he distinguishes, ought to augment the standard principles of bioethics and influence ethical decision-making to focus on the vulnerable. The poor, elderly, young, or disenfranchised are often made more

vulnerable by certain social conditions. Solidarity as a principle to aid the vulnerable in this sense demands social measures to ameliorate the vulnerability and strive for equality. Acting on this principle may include welfare policies to try to create more just and equitable conditions, but it may also mean that the vulnerable populations ought to be brought into consideration or even be the focus of any public policy (Stone 2012).

Sass also suggests that "solidarity" may be understood as "justice" within the context of principles of bioethics. Further solidarity is often paired with a principle of subsidiarity in European tradition. Subsidiarity means that decisions ought to be made within the community that has the relevant information and that is most directly affected by the decision. Hence, decisions that affect the family ought to be made by the family if it has the requisite information. Higher levels of decision-making bodies enter when the requisite information is not present. Solidarity and subsidiarity work together because the solidaristic community provides both the context for the decision and the information on which the outcome is based.

An alternative approach to the principle of solidarity is that the principle of solidarity supports and sustains other principles operative in the bioethics literature such as justice and benevolence. Solidarity, in other words, may be a condition for acting on moral principles. This approach is sometimes seen in human rights discussions (Harmon 2006; Gunson 2009; Scholz 2014); a willingness to aid others and a recognition of social bonds are, it seems, a presumption or requirement for responding to human rights principles. Solidarity points to the social bonds and institutional mechanisms necessary for justice.

Solidarity for Global Bioethics

Solidarity as a moral relation and as a principle has come to play a prominent role in bioethics and biolaw literature globally. Five specific contexts for the discussion of solidarity are public health, social security and the welfare state, universal health care, the right to health care, and aid to the poor and socially vulnerable (see also Prainsack and Buyx 2012).

"Solidarity" in bioethics applies the insights of the moral relations and principles of solidarity to access to health care within a state and beyond state boundaries. In large part, state-based policies that ensure access to health care and prevention of disease emerge out of a larger commitment to welfare for citizens. The connection between health and welfare may be justified through the fact that the well-being of individuals serves the common good or through an obligation to aid the sick and infirm (see van Donselaar 1998). Consumerism, individualism, and privatization pull against solidarity in the distribution of health care (Stjernø 2005, p. 338; Baylis et al. 2008, p. 204).

Public Health

Public health is generally understood as focusing on the common good by maintaining the formal organization of institutions and practices with the goal of preventing disease, promoting health, and coordinating care or access to health provisions. Civic solidarity, instantiating the state's obligations to protect citizens against social vulnerabilities, includes or requires significant attention to health-related aspects of social life. Domestically, social life creates multiple avenues that may make some citizens more vulnerable than others. The placement of low-incoming housing in close proximity to polluting factories, for instance, exacerbates already existing material inequalities thereby making this subset of the population more vulnerable. Similarly, the lack of access to routine medical exams and the cost of tools for proper hygiene potentially affect persons in the lower rungs of the income brackets more than others. These public health issues, as well as new threats due to relations of globalization and terrorism, may also be seen on a global scale. Solidarity in the context of public health focuses on measures that ensure communal well-being; it may be grounded in mutual recognition (Houtepen and ter Meulen 2000a) or on more inclusive accounts of interdependency or interconnection (Baylis et al. 2008). In either case, solidarity shapes public policy to counter excessive individualism and

ensure equal access to the means for preventing disease, accessing health care, and promoting flourishing life.

Relations of solidarity are particularly important in global public health debates; the threat of pandemics requires a global perspective toward public health. Solidarity can play a central role in an ethical framework that seeks to justify obligations across borders as well as protect especially the most vulnerable populations. The ease and pace of travel further create a situation wherein diseases can more quickly spread across borders and around the world. Epidemics can no longer be thought of as contained and containable within nations or even regions. Coordination among nations and within the international community is necessary in order to meet these new challenges. Baylis, Kenny, and Sherwin (2008), for instance, describe a relational solidarity, paired with relational personhood, specifically aimed at public health planning for pandemics. Solidarity is the umbrella concept for international coordination of this sort.

Social Security and Welfare State

Solidarity in the European tradition is instantiated in and synonymous with the welfare state. Welfare systems vary but the basic premise is that resources are reallocated to provide for the well-being of all citizens. One way to do this is through a well-established insurance system and social security in the event of disability or retirement. Solidarity seeks to ensure that hard times may be weathered more easily by ensuring that social networks can maintain the basic needs of all citizens. Because it is a universal good, health care generally has more public support, even when support for other welfare programs declines (Houtepen and ter Meulen 2000b; Bergmark 2000). Health care and other welfare programs ensure political inclusion by decreasing the social vulnerabilities that might inhibit an individual's ability to participate in civil and political life.

One of the many debates affecting conceptions of solidarity as welfare is whether the achievement of equality in health-care provision evinces a high degree of solidarity or whether extensive state welfare regimes for the distribution of health care and elder care actually erode solidarity networks like families, neighborhoods, and local communities. This is referred to as the debate between the value of institutionalized solidarity and the importance of a solidarity that emerges through individual social bonds or connections to others that inspire what are called individual expressions of solidarity. Another way to think of this is as a debate between civic solidarity (or the obligations of the state to its citizens) and social solidarity (or the cohesion between fellow citizens and their willingness to respond to each other's needs).

Addressing the situation in Europe, Houtepen and ter Meulen (2000b) suggest that within some contexts, the traditional form of solidarity, rooted in voluntary societies, has given way to institutionalized forms of risk sharing they call "contractual solidarity in welfare." One potentiality, however, is that the demand for health care will continue to outpace the supply – a growing concern given increasing life expectancies – and intergenerational solidarity will continue to erode. This example illustrates the practical importance of the debate regarding institutionalized solidarity versus individual expressions of solidarity.

Universal Health Care

Arguments for universal health care appeal to solidarity both as a principled justification for universal provisions and as a relational commitment to universal provision. Universal health care means that health care ought to be universally available and that the quality of care ought to be equitably distributed. Differences in socioeconomic status ought not to affect the accessibility of care; of course, public health policies seek to ameliorate inequalities within individual nations, but international organizations are charged with the responsibility to ensure equitable global distribution of health resources. Thus a mandate to attend to the needs of poorer nations is incorporated in the calls for universal health care.

The global disparity in access to health care parallels or is causally related to global economic

inequality. Solidarity in this context means there is a communal obligation to address health-care needs, especially for the most vulnerable, and some might also suggest that there needs to be more responsibility on the part of international organizations to pressure nations to provide universal access to health care and social security in the case of age, illness, or disability. Of course, it might also be argued that justice, or more specifically distributive justice, is more appropriate to the discussion of universalized health care (Butler 2012), but such a claim echoes the debate regarding the relation between justice and solidarity discussed earlier.

Right to Health Care

Another meaning of solidarity in bioethics concerns a right to health care, a subsidiary of the human right to health. Although in some ways claiming a right to health is impossible insofar as health is simply out of reach for some individuals, the right implies that every human being has a right to the resources necessary for the highest possible attainment of health. These include the systems discussed under public health and universal health care, i.e., access to medical care, social security, healthy living conditions (proper sanitation, adequate food, and safe environment). Solidarity functions as both a relation and a principle in this context. Recognition of human interdependence – including the acknowledgment of how pollution in one part of the world adversely affects other regions – informs a network of moral obligations that both avoid the interference in the right to health and work to positively ensure that right for all others. The former, noninterference in the right to health, is justice; the latter, positive action to ensure the exercise of that right, is solidarity.

As a principle, solidarity stands in for the right to health in a nonreductive way. That is, the right to health is really a complex right operative on interpersonal, national, and international levels and involving all those aspects of life and social existence that impact mental, physical, and material well-being. The principle of solidarity stands as an indicator of that inclusive understanding of the right to health; it also challenges moral theorists to incorporate the interconnections of peoples and actions in the shaping of theory and policy.

The United Nations specifies its mandate to promote solutions to health problems internationally in its Charter, Chapter IX, Article 55:

> With a view to the creation of conditions of stability and well-being which are necessary for peaceful and friendly relations among nations based on respect for the principle of equal rights and self-determination, the United Nations shall promote: (a) higher standards of living, full employment, and conditions of economic and social progress and development; (b) solutions of international economic, social, health, and related problems; and international cultural and educational cooperation; and (c) universal respect for, and observance of, human rights and fundamental freedoms for all without distinction as to race, sex, language, or religion.

Subsequent documents further elaborate on the right to health. Gunson (2009) shows how the Universal Declaration on Bioethics and Human Rights, although mentioning solidarity only three times in the text, is actually built on a framework of the principle of solidarity. The same may be said for other agreements and statements of the international community.

Similarly, the Charter of Fundamental Rights of the European Union includes the right to preventive health care (Article 35; EUROPA 2000). Basic preventative health care is a good for society as a whole; it has a positive effect on well-being individually and communally, tends to decrease emergency health expenditures, and plays a significant role in planning to avoid transnational pandemics.

Focus on the Vulnerable

Finally, throughout all discussions of solidarity is a strong thread, already alluded to, that mandates a focus on the poor and vulnerable. It is this sense that is evident in the Catholic social teaching and a wide variety of social movements. The former calls for a dedicated commitment to remedying the inequalities that keep fellow human beings from full human dignity.

As explained previously, one way to understand the principle of solidarity is as an

augmentation to traditional principles of bioethics that redirects the focus of concern to the most vulnerable. Motivating that focus may be accomplished through a recognition of ourselves in those who are vulnerable or as a rationally justified obligation to aid those in need. There is a long tradition using the term "solidarity" to indicate the need for this focus on the poor and vulnerable.

Conclusion

Although each of the categories of solidarity within bioethics may be analyzed distinctly, there is significant overlap as well. A right to health care may be achieved through the provision of universal health care; health care as an element of welfare provisions helps to meet the needs of vulnerable populations. Solidarity manifests so differently because of the varying nature of social bonds as well as the aims or goals of the solidaristic community. All forms share in common the willingness to acknowledge social bonds with others and to act on behalf of the needs or interests of those others. It is also precisely this that leads some to wonder whether there is a crisis of solidarity. Individualism and consumerism have infected social existence to such a degree that even those societies that have supported and defended health care as a requirement of solidarity are beginning to question that commitment (Houtepen and ter Meulen 2000b).

Cross-References

- Access to Healthcare
- Common Good
- Empathy
- Human Dignity
- Human Rights
- Justice: Global
- Justice: Theories of
- Public Health
- Responsibility: Collective
- Social Ethics
- Vulnerability
- Welfare

References

Bayertz, K. (1999). Four uses of "solidarity". In K. Bayertz (Ed.), *Solidarity* (pp. 3–28). Dordrecht: Kluwer.

Baylis, F., Kenny, N. P., & Sherwin, S. (2008). A relational account of public health ethics. *Public Health Ethics, 1*(3), 196–209.

Bergmark, Å. (2000). Solidarity in Swedish welfare: Standing the test of time? *Health Care Analysis, 8*(4), 395–411.

Butler, S. (2012). A dialectic of cooperation and competition: Solidarity and universal health care provision. *Bioethics, 26*(7), 351–360.

EUROPA. (2000). European Commission: The charter of the fundamental rights of the European Union. http://ec.europa.eu/justice/fundamental-rights/charter/index_en.htm. Accessed 15 Nov 2015.

Gunson, D. (2009). Solidarity and the universal declaration on bioethics and human rights. *Journal of Medicine and Philosophy, 34*(3), 241–260.

Harmon, S. H. E. (2006). Solidarity: A (new) ethic for global health policy. *Health Care Analysis, 14*(4), 215–236.

Hoedemaekers, R., & Dekkers, W. (2003). Justice and solidarity in priority setting in health care. *Health Care Analysis, 11*(4), 325–334.

Houtepen, R., & ter Meulen, R. (2000a). The expectation (s) of solidarity: Matters of justice, responsibility, and identity in the reconstruction of the health care system. *Health Care Analysis, 8*(4), 355–376.

Houtepen, R., & ter Meulen, R. (2000b). New types of solidarity in the European welfare state. *Health Care Analysis, 8*(4), 329–340.

Prainsack, B., & Buyx, A. (2012). Solidarity in contemporary bioethics: Towards a new approach. *Bioethics, 26*(7), 343–350.

Sass, H.-M. (2001). Introduction: European bioethics on a rocky road. *Journal of Medicine and Philosophy, 26*(3), 215–224.

Scholz, S. (2008). *Political solidarity*. State College: Pennsylvania State University Press.

Scholz, S. (2014). Solidarity as a human right. *Archiv des Völkerrechts, 52*(1), 49–67.

Stjernø, S. (2005). *Solidarity in Europe: The history of an idea*. Cambridge: Cambridge University Press.

Stone, J. (2012). Elderly and older racial/ethnic minority healthcare inequalities: Care, solidarity, and action. *Cambridge Quarterly of Healthcare Ethics, 21*, 342–352.

Van Donselaar, G. (1998). The freedom-based account of solidarity and basic income. *Ethical Theory and Moral Practice, 1*, 313–333.

Further Readings

Brunkhorst, H. (2005). *Solidarity: From civic friendship to a global legal community*. Cambridge, MA: MIT Press.

Laitinen, A., & Pessi, A. B. (2015). *Solidarity: Theory and practice*. New York: Lexington Books.

Ter Meulen, R. (Ed.). (2001). *Solidarity in health and social care in Europe*. Dordrecht: Kluwer.

Spirituality

Carl Becker
Kokoro Research Center, Kyoto University, Kyoto, Japan

Abstract

Spirituality refers to the search for meaning and understanding of life, with reference to nonphysical values or powers. The growing recognition of spirituality and its effects on medical outcomes leads to ethical questions including (A) proper assessment of and response to spirituality in medical situations, (B) what to do when someone's spirituality contradicts standard biomedical judgment, and (C) the underlying assumptions of bioethics as typically presented in the West. These are questions of: how far assessing (or ignoring) patients' spirituality would be ethical, how far accepting (or overriding) patients' spiritually grounded decisions would be ethical, and whether the presuppositions of modern Western bioethics ultimately conflict with the presuppositions of spirituality.

Keywords

Spirituality; Assessment; Medical ethics; Religion; Worldviews

Introduction

Spirituality is a term used to cover both formal religion and personal values and beliefs about the relation of consciousness to invisible dimensions "that impart vitality and meaning to life's events" (Maugans 1996). Disaster, disease, and death challenge our sense of justice and reason; personal crises, such as divorce, unemployment, incapacitation, dying, or bereavement, lead people to ask "why me?" "how can I make sense of this?" "what is life/the universe trying to teach me?" and "what happens hereafter?" Scholars often distinguish existential elements (wonder, awe, intuited harmony, inspiration or ecstasy, commitment to a search for truth) and transcendent elements (karma and rebirth, judgment and afterlife, connection with a higher power) within spirituality. Concern with spirituality builds on Allport's (1964) classical distinctions between extrinsic and intrinsic religion, observing that while some churchgoers apparently lack spiritual concerns, even agnostics with no religious affiliation may live by strong beliefs and commitments.

Spirituality has important implications for bioethical issues concerning war, capital punishment, and the environment, but this entry focuses on medical and psychological bioethics. Formal religions take positions on bioethical issues ranging from abortion and euthanasia to circumcision and blood transfusion. The present entry focuses on bioethical issues arising *outside* of formal religious belief systems but within the broader search for human meaning and interconnection known as spirituality. Concerns with spirituality raise bioethical issues on three levels, viz., (A) assessment, (B) biomedical judgment, and (C) challenging of bioethical assumptions.

History and Background

As personal worldviews have supplanted the role of community faith in secularizing societies, the term "spirituality" has largely replaced the term "religious." In the 1970s, American public opinion polls exposed significant declines in the numbers of people willing to call themselves "religious," despite an unmistakable upswing in "New Spirituality" and "New Age" movements. When public opinion polls asked respondents about "spiritual" (rather than religious) interests or belief in "invisible spiritual forces" (rather than God), overwhelming affirmative responses yielded results far more congruent with the sensed tenor of the times. In the 1980s, the United Nations World Health Organization began to use the term spirituality along with religion in its discussions of health and quality of life, and by the 1990s, the WHO considered the term spirituality as an important dimension of health. While communist delegates urged adding the proviso "in some countries" or broadening spirituality to

include "attitudes and philosophies," subsequent discussions have centered more on how to put such concerns into practice.

Outside of Western medicine, care for sufferers' spiritual needs typically has preceded and predominated over physiological care; in recent decades, even Western hospices and hospitals increasingly recognize the spiritual needs of patients and bereaved. Merely pharmaceutical response to the symptoms of an illness, injury, or bereavement, without addressing their underlying meanings, is like bombing cities in response to suicidal terrorists; ignoring the fundamental causes of disharmony, it exacerbates malaise and alienation rather than attaining long-term solutions.

Responding to patients' spiritual needs can also improve medical care, by referring patients to chaplains or counselors, using spiritual support resources in the community, incorporating therapeutic touch and "healing" music or aromas, and teaching meditation or relaxation for chronic pain or insomnia. Although debates continue about the validity of therapeutic touch, healing prayer, and meditation, these spiritual responses to medical issues have become objects of formal research rather than of peremptory dismissal, and often complement medical procedures. Yet the areas of spirituality most subject to bioethical discussion are not the utility of prayer and meditation but how patient spirituality should be assessed and valued, especially when it influences medical decision-making.

Assessment

Reasons for Assessment

Hospital ethics conferences use criteria like best interests, benefits/burdens, performance, QALYs, and DALYs to ground medical decisions, but for many patients and families, the more central questions concern the meaning of suffering: whether they can find redemption, forgiveness, or higher purpose amid tragedy. These are questions of spirituality. Scholarly studies suggest that subjects' spirituality not only reduces suicides and unethical behaviors but also contributes to psychological well-being and physical health, including recuperation from disability, illness, and loss (Pargament 2011; Koenig 2009).

Feeling a sense of purpose in rehabilitation or of meaning in cancer treatment can significantly improve patients' compliance and speed their recovery. Conversely, belief that suffering is good or karmically deserved can impede healthy outcomes. If doctors have an ethical duty to diagnose the causes and factors affecting their patients' health, are they similarly obligated to assess their patients' psycho-spiritual concerns that affect health outcomes and to become competent to do so?

Doctors often prescribe futile medicines and procedures for disabling or terminal illness, more for the sake of raising patients' and families' hopes than for their demonstrable cost-effectiveness. If it were known that patients and families would feel greater hope through ritual, meditation, music, prayer, sacred narratives, or inspirational readings, is it ethical to withhold those forms of spiritual care, when their side effects were less harmful than a predictably futile round of chemotherapy? The uncertainties inherent in such diagnoses and prognoses should not ethically change the patients' set of choices.

History of Assessment

Some doctors still think that tools for psycho-spiritual assessment lack the level of precision required in medicine or that diagnostic data is too weak. If the issue were the precision of medical prediction, doctors themselves would face an unbearable burden of proof. If the issue is the accuracy not of prediction but of diagnosis, a wide range of spiritual diagnostic assessment tools are demonstrably useful in treating patients.

As early as 1982, Paloutzian and Ellison's SWBS (*Spiritual Well-Being Scale*) (Paloutzian and Park 2013) assessed religious well-being relating to God (e.g., "God loves me and cares about me") and existential well-being relating to purpose in life (e.g., "life is full of conflict and unhappiness"). The SWBS has since been criticized for being too monotheistic and for failing to distinguish variations among people with high spirituality. Moberg (2001) advanced the SWBS

to a 94-item questionnaire, but this proved too tedious for clinical settings.

By the late 1990s, physicians like Maugans advocated taking spiritual histories of their patients (1996); Harvard University (Puchalski 2006), and the University of Hull's Centre for Spirituality Studies (McSherry and Ross 2010) validated reliable protocols for doing so. In the 2000s, the FACIT organization developed the Functional Assessment of Chronic Illness Therapy – Spiritual Well-Being (FACIT-Sp 2014) that measures sense of meaning and the role of faith in illness, with items like "I have trouble feeling peace of mind" and "my illness has strengthened my faith or spiritual beliefs" (http://www.facit.org/FACITOrg/Questionnaires).

Of the dozens of spiritual assessment tools now circulated and taught, the FACIT-Sp and SWBS have become most widely used. The American Association of Medical Colleges (AAMC) has developed medical school objectives (MSO) related to spirituality and culture. So the issue has evolved from how spirituality can be measured to whether it should be measured proactively.

Ethical Issues in Assessment

Can physicians and hospitals better respect and support their patients' beliefs by ignoring them or by proactively probing them? Are they ethically required not only to respect but also to support their patients' spirituality? Doctors maintain they are not ethically bound to pray for nor discuss karma with their patients – but are they ethically required to seek someone who will do so, if this would substantially affect their patients' outcome? If patients ask for spiritual support, can doctors ethically deny it? Can doctors ethically prohibit praying or chanting in hospital wards? What should be done when the values of medical counseling seem to obscure or conflict with the patients' spirituality?

Medicine tries to treat the physical aspects of pain, disease, and suffering while ignoring their important spiritual aspects. Many doctors feel ethically obliged to require and disclose the results of diagnostic tests even to patients who desire neither testing nor information. If spirituality is a major factor in health, are doctors ethically obliged to require spirituality tests, even to patients who do not desire such tests or information? (If doctors feel unprepared or uncomfortable discussing spirituality, this is a reason not to avoid such testing but to improve doctors' preparedness.)

Religious hospitals that refuse to provide abortions should seem ethically required to inform patients of the effects of not aborting and their choices of other hospitals that will provide abortions. Then should materialist hospitals that refuse to provide spiritual assessment or counseling be ethically required to inform patients of the effects of ignoring spiritual assessment and their choices of other hospitals that will provide spiritual assessment or counseling?

Responses to these questions about ethical requirement of spiritual assessment range from "always" to "never." To reduce the risk of intrusion and offense in directly asking *all* patients about their beliefs and commitments, many practitioners now suggest prescreening *before* spiritual assessment. In other words, before asking a patient specifically about beliefs or faith community, a prior question should be asked, "are spiritual or religious issues important to you?" or "are there any spiritual concerns you would like to have someone help you address?"

The preponderance favors assessment (often deferred to chaplains or psychologists) only when patients proactively express desire for spiritual support or counseling (McSherry and Ross 2010). Yet this may overlook many patients who harbor the same spiritual needs but hesitate to voice them in medical settings, not to mention the possibility that medical professionals themselves might benefit from spiritual care. If further evidence shows that patients and bereaved families receiving spiritual care are less devastated by grief and less likely to sue, economic reasons alone may promote the assessment and treatment of spiritual crises.

The question of universal spiritual assessment is somewhat analogous to that of universal cholesterol assessment and prescription of statins. Just as there is no precise cholesterol count indicative of statin treatment nor precise way of

predicting the statins' side effects, there is no precise level of anxiety indicative of spiritual counseling nor precise way of predicting the side effects of the counseling. In the case of cholesterol and statins, however, the debate has focused largely on the economics of screening versus the benefits of medication. In the case of spiritual pain, the focus extends to the rights and worldviews of the patients themselves – a more quintessentially ethical area.

Further spiritual dilemmas remain. Are physicians obliged to treat patients whose worldviews utterly contradict their own – like suicide attempters or terrorists who promise to repeat their attempts if their present physicians restore them? Physicians already attempt to dissuade patients from delusions about prognoses, on the grounds of their superior medical knowledge. Then should physicians also attempt to dissuade patients from what the physicians think are delusions about God's will or the futility of life – and on what grounds?

Ethical issues can arise when doctors or patients blur medical and spiritual roles which society wants separated. The ethics of universal spiritual assessment focus on benefiting and not harming the patient, as medicine is increasingly "patient centered." Yet the very notion of "patient-centered" medicine itself raises a second level of ethical questions.

Biomedical Judgment: Ethics of Decision-Making

When physicians are faced with decisions about the desirability of CPR or aggressive life support, they rely on their beliefs about the meaning and value of human life or their understanding of their commitments as physicians. These are spiritual beliefs. Lacking medical knowledge, patients rely even more heavily on their beliefs about the meaning and value of human life or their understanding of their commitments to family or to God. So personal spirituality may have important implications for stances on bioethical issues. Even nonreligious physicians should not underestimate the effects of their patients' belief systems. Yet the absence of formal denominational labels raises problems, not only in assessment but also in response to patients whose worldviews differ.

Spiritual commitments can underlie such statements as:

I want to bear my babies at home.
I will not pollute my body with inorganic chemicals or animal products.
I do not want to give my beating heart or liver to someone else.
If I can no longer eat nor communicate, I do not want my body prolonged.
My suffering here and now will burn off my previous bad karma.
I must spend my last moments lying on the ground (or floor), not in a bed.
I want to be frozen until medicine can revive me and cure this disease.
I want no one to try to revive me, lest my soul be torn between two realms.
I want everything possible done for me because the universe will save me with a miracle.

Obviously, such positions may be more or less rational, more or less negotiable. They may be based on traditionally sectarian viewpoints or on personal spirituality. To measure their depth and strength requires not only standard spiritual assessment tools but a close understanding of the patients' lifestyle hitherto.

Whether a belief is spiritual or not, if it is almost certainly mistaken (such as "the universe will save me with a miracle" or "unfreeze and revive me after my disease becomes curable"), then physicians have the difficult task of helping people overcome such delusions while respecting them as persons (cf. Cochrane 2007). Conversely, when a patient will endure suffering or resist organ donation for spiritual reasons, it may help medical practitioners to understand the spiritual source of that resistance. In some cases, a sympathetic understanding may open the door for a deeper dialogue about what the patient thinks God or the universe wants and ultimately to a reframing or reinterpretation that allows other treatments without requiring a conversion of worldviews. The desires of Hindu patients to die on the floor, or of Tibetans not to be touched for

some time after breathing stops, initially challenge hospital procedures, but cultural sensitivity can accommodate such practices if the medical system is adequately forewarned of the patient's spiritual desires.

Thornier ethical questions concern how to deal with spirituality that conflicts with medical judgment. When a Jain or Quaker claims that their religion requires conscientious objection to military conscription, when a Jehovah's Witness or Christian Scientist rejects blood transfusions, when a Catholic or Muslim refuses to abort a deformed fetus, their religious affiliation is prima facie evidence of their commitment. Their faith community may legally and socially support their violation of the majority ethical opinion, even calling into question the ethicality of transfusion or abortion itself.

However, when someone with no religious affiliation claims that her spiritual worldview requires physician-assisted suicide, the grounds for this claim are more difficult to demonstrate. How can a physician know that a patient really understands her situation and is deciding not from ignorance and fear but from spiritual commitment? Many would suggest that it is not a physician's job to distinguish fear from spiritual commitment, much less to attempt to influence his patients' spiritual commitment. But when a spiritual belief system is cited as a reason for choosing or rejecting courses of treatment in ways the physician thinks suboptimal if not dangerous, then is the physician obliged to follow the patient's wishes or to introduce someone who will? Some countries (like America) tend to prioritize the personal rights of the patient over that of the trained medical worldview; others prioritize trained medical opinion over that of the layperson (as in Britain's ***Bolam v Friern 1957***).

Choice of life-prolonging treatment (among many other ethical choices like those above) is not like a choice of cream or sugar in tea or coffee; it involves our very understanding of life and the world. Since spirituality is a commitment to a particular personal understanding of life, patients' spirituality may have important implications for their personal stances on bioethical issues. How far must law or society ethically respect the spiritual positions of others, when those positions affect life, death, or the use of common medical resources? Neither courts nor bioethicists have reached clear conclusions on such issues, but as modern people move away from organized religions toward increasingly personal spiritual belief systems, both courts and bioethicists will be increasingly forced to address these questions, particularly in situations where patients' preferences disagree with their doctors'. Conversely, whereas hospital entrance forms could traditionally ask patients' religious affiliation, future understanding of patients' spiritual beliefs will require ethically sensitive measures to identify the reasons behind patients' treatment decisions – and perhaps to provide counselors or chaplains who can discuss those decisions on spiritual as well as medical grounds.

Challenging Bioethical Assumptions

If spirituality were reducible to existential angst, then physicians might be allowed ethically to override the ignorant fears of their patients. On the other hand, if spirituality highlights cross-cultural relativity and transcendent values, then it ultimately challenges the assumptions, not merely of medicine but even of bioethics as it has evolved in the West. Western bioethics typically accepts the principles of autonomy, beneficence, and justice as self-evident. Many more spiritual perspectives challenge the cultural and economic biases underlying the proselytization and use of these very "principles."

Autonomy

The "principle" of "autonomy" is as recent as it is biased. The notion arose from the eighteenth-century white property-owning male enlightenment thinkers who analogized the universe to a clock and life to deducing mathematical theories from principles. They never imagined rights of women, servants, children, or other races, much less of societies or animals.

Many spiritually minded cultures find the notion of autonomy incomprehensible. The very word autonomy is untranslatable and therefore

unthinkable in many cultures and languages, so to impose it on them is little less than bioethical imperialism. In other societies that have learned to translate the term "autonomy," it exemplifies *un*ethical rather than ethical thinking. Traditional societies from China and Japan to Polynesia and sub-Saharan Africa hold that humanhood is quintessentially relational and inextricably social. For many educated Asians and Africans, an ethical decision is one that considers all the impacts and desires of all the people and groups that might have preferences about it. Someone who decides *for* themselves, *about* themselves, and *by* themselves is at best lacking in social maturity and consideration and at worst criminally self-centered. In such cultures, intuitive understandings of rightness based on human interactions and concern for future generations are felt to be far more ethical than principled calculations or signatures on incomprehensible consent forms.

Beneficence

The bioethical "principle" of "beneficence" tends to prioritize short-term benefit, based on limited knowledge about long-term side effects and societal and environmental impacts. Thus, statins may provide a short-term fix for high cholesterol, but in the long run, their side effects may override their benefits, where changes of patient diet and lifestyle would prove far preferable for the patient. Steroids or antibiotics that seem beneficial in the short run may in the long run create allergies or antibiotic-resistant pathogens that threaten society. Transplanting a resected liver into a waiting recipient may benefit the recipient in the short run but may risk reducing the quantity or quality of life of the donor in the long run – and possibly lead to social commodification of human body parts. To provide such drugs or organs in ways that may threaten the physical health or ethical thinking of society in the future is unthinkably unethical from some spiritual perspectives.

Many spiritual worldviews maintain that long-term sustainability is ethically preferable to quick fixes. Many traditional cultures consider ethical effects of actions on many generations of descendants, on society, and on environments centuries hence. Their spirituality would resist using recently discovered chemicals in human foods and medications before their safety and side effects had been proven on generations of unwitting human guinea pigs. It would reject meat eating for its deleterious impacts both on environment and on human health, not to mention on the sentient animals which are butchered. Traditional spirituality would be extremely cautious about genetic engineering with the potential to disrupt many delicate balances in nature, even under the guise of "beneficence."

Justice

The bioethical "principle" of "justice" or "fairness" tends to prioritize supposed equality of choice over equality of outcome and material quantity over psycho-spiritual quality. Modern capitalists tend to equate well-being with economic prosperity, if not reducing happiness to material wealth or longevity. Similarly, medical notions of fairness all too often look at dollars spent or length of patient survival, without considering the psycho-spiritual quality of the life of the bedridden or even unconscious patient. Many spiritual perspectives would laud the patient who wishes a shorter conscious ambulatory life over a longer unconscious bedridden one. Yet bioethical notions of fairness all too often refer implicitly to economics – the money and resources to be spent on a given patient – rather than on trying to enable patients to reach sense of meaning and peace within their situations or greater satisfaction in their deaths.

Fox and Swazey (2008) have cogently critiqued the hegemonic thrust and cultural myopia of American bioethics and its failure to address issues of international injustices and inequities in health care – partly marginalized by bioethicists' unanalyzed tendencies to consider them as economic *rather than* ethical problems. For example, under the rubric of "justice" or "fairness," Western bioethicists glibly debate the ethics of providing costly life-extending liver transplantation to a minuscule minority of wealthy Westerners while ignoring millions of children suffering organ damage and physical handicaps as a result of malnutrition. Recent debates over the desirability of costly genetic enhancement of embryos beyond any natural norm take place against an unseen

background of countless babies born with crippling genetic defects that will never be addressed by their societies' medical resources. Bioethicists debate the use of embryonic stem cells to enable infertile or homosexual couples to replicate themselves, while tens of thousands of orphans in need of good homes cry out for adoption. A more spiritual view looks not only at fairness for individuals but at fairness for a larger and longer humankind. It suggests that, while not all unfairness can be addressed, as long as a vast portion of the world lacks elementary medical care and hygiene, debates on the ethics of costly advanced medical techniques are cruel and inconsiderate at best and at worst make a travesty of any pretense to "fairness" or "justice."

Ethics, not Calculus

Finally, the presupposition that ethics should be deduced from culture-blind or culture-transcendent principles directly violates spiritual insights that ethics arise from human emotional interactions within concrete cultural situations. Since spirituality is not a single position or denomination, it cannot be said that all spiritualities oppose such principlism. But spirituality implies a plea to consider less what scalpels or chemicals may do to patients' cells and more what interactions or power relations may do to people's souls. Biomedical decisions should not be made from the falsifiable fiction of a just and rational world operating according to Cartesian principles; they must address social, cultural, and long-term implications of treating people as bodies or intelligent objects, rather than as hearts, souls, and subjects in search of meaning in this life and perhaps the next. From this perspective, spiritual crises challenge not only particular medical assessments and procedures but indeed the very presuppositions underlying the unconsciously hegemonic movement of Anglo-European bioethics into traditional Asian and African worldviews.

Conclusion

Times of crisis require insights into their existential implications. Incurable cancer and dementia sufferers – and their caregivers or bereaved families – demand more than X-rays and drugs; they want to understand not only the causes but also the potential meanings of their tragedies. Spirituality implies prioritizing people's feelings, beliefs, and meanings over material measurements of cells, organs, or life expectancies. Taking spirituality seriously requires not only listening to patients but also treating their worldviews with respect. Many ethical issues linger concerning the ways and extents of respecting spirituality. Ultimately, concerns with overarching meaning and transcendence may challenge not only specific medical procedures and prejudices but the very grounds and scope of twentieth-century Western biomedical ethics itself.

Acknowledgments This entry is the product of an employee of (National) Kyoto University, Japan, but the statements, opinions, or conclusions contained herein do not represent the statements, opinions, or conclusions of Kyoto University or the Japanese government.

Cross-References

- ▶ Care Ethics
- ▶ Confidentiality
- ▶ Life: Meaning of
- ▶ Medicine and Ethics
- ▶ Psychiatry Ethics
- ▶ Religion and Global Bioethics

References

Allport, G. W. (1964). *The Individual and His Religion: A Psychological Interpretation*. New York: Macmillan, Allport.

Cochrane, T. I. (2007). Religious delusions and the limits of spirituality in decision-making. *American Journal of Bioethics, 7*(7), 15.

FACIT-Sp. (2014). http://www.facit.org/FACITOrg/Questionnaires

Fox, R. C., & Swazey, J. P. (2008). *Observing bioethics*. Oxford: Oxford University Press.

Koenig, H. (2009). *Spirituality in patient care: Why, how, when, and what* (2nd ed.). Philadelphia: Templeton Press.

Maugans, T. A. (1996). The SPIRITual history. *Archives of Family Medicine, 5*, 11–16.

McSherry, W., & Ross, L. (2010). *Spiritual assessment in healthcare practice*. Cumbria: M&K Update.
Moberg, D. O. (2001). *Aging and spirituality: Spiritual dimensions of aging theory*. New York: Haworth.
Paloutzian, R. F., & Park, C. L. (Eds.). (2013). *Handbook of the psychology of religion and spirituality* (2nd ed.). New York: Guilford Press.
Pargament, K. I. (2011). *Spiritually integrated psychotherapy: Understanding and addressing the sacred*. New York: Guilford Press.
Puchalski, C. M. (2006). *A time for listening and caring: Spirituality and the care of the chronically ill and dying*. Oxford: Oxford University Press.

Further Readings

Cobb, M., Puchalski, C. M., & Rumbold, B. (Eds.). (2012). *Oxford textbook of spirituality in healthcare*. Oxford: Oxford University Press.
Emmons, R. A. (1999). *The psychology of ultimate concerns: Motivation and spirituality in personality*. New York: Guilford Press.
Johnston-Taylor, E. (2007). *What do I say? Talking with patients about spirituality*. Philadelphia: Templeton Press.
Lazenby, M., McCorkle, R., & Sulmasy, D. P. (Eds.). (2014). *Safe passage: A global spiritual sourcebook for care at the end of life*. Oxford: Oxford University Press.

Sports

Andy Miah
School of Environment and Life Sciences, University of Salford, Manchester, UK

Abstract

This chapter explores the relationship between bioethics and sport ethics, which changed dramatically in the early 2000 when the genetics era generated a series of new questions about the ends of sport and how they would interface more widely with a range of bioethical principles. Focused largely on the nontherapeutic application of genetics to persons, the entry situates these debates within the context of discussions about the use of human enhancement and wider debates about transhumanism. It argues that concerns about the ethics of performance enhancement in sport have become more closely aligned with wider public health concerns, where doping should be seen as more than just a problem for elite sport. It also examines the overlap between technologies, which have further expanded the field of bioethics into such areas as disability studies, where the case of Oscar Pistorius – as the first prosthetically enabled Paralympian to compete within the Olympic Games – has been a prominent example of the overlap between biotechnology and biomechanical prosthesis.

Keywords

Sport; Gene doping; Human enhancement; Transhumanism; Posthumanism; Prosthetics

Introduction

The applied ethical traditions of bioethics and sport ethics grew up in the 1970s, but largely without reference to one another. The foundation of the Philosophic Society for the Study of Sport (now the International Association for the Philosophy of Sport) by Paul Weiss in 1972 generated interest particularly around the ethical issues related to doping, which became a prominent topic throughout the 1980s. In part, this may be attributed to a number of high-profile cases of athletes being caught using doping methods or substances, as in the case of Canada's sprinter Ben Johnson, which revealed a culture of doping that was rife in athletics.

The social concern about the prominence of doping in sport, as a threat to crucial values and the image of sport, were important catalysts for debates about the ethics of doping among sport philosophers, where scholars interrogated the prima facie assumption that doping was wrong by drawing attention to the peculiar practices of sports, which operate with rules that are different from those that are implemented within society at large. Yet, even here, where the topic involved the application of medical substances to, arguably, nonmedical use, the points of connection between sport ethicists and bioethicists were few and far between. Sports ethicists – who came largely from

kinesiology and physical education departments, with some roots in the philosophy of education or phenomenology, focused on the concepts of cheating, fair play, the body, and justice, arguing that sport provides unique examples where these ideas do not fit with traditional interpretations.

Meanwhile, bioethicists focused on the ethics of science and medicine to approach the same subject. Around this early period, bioethics emerged as a distinct area of inquiry from medical ethics, with such institutions as The Hastings Center establishing new territory and expanding the remit of ethicists within scientific and medical research and practice. Indeed, The Hasting Center was a crucial point of interface within the history of bioethics and sport. In 1983, Thomas Murray, who would later become a key figure in the World Anti-Doping Agency's ethical issues review panel, the US Anti-Doping Agency work, and President of The Hastings Center, described the "coercive power of drugs in sport," arguing that "the use of performance enhancing drugs is ethically undesirable because it is coercive, has significant potential for harm, and advances no social value" (Murray 1983, p. 30).

Soon after, Norman Fost (1986) developed his "skeptical view" on banning drugs from sport, but again, operating largely outside of the philosophy of sport community. Around this period, Murray was also more broadly active in articulating a number of other bioethical tensions within sports practice, such as the divided loyalties of an athletic team physician, who is under an ethical obligation to ensure the health of players is protected, while also will find themselves expected to make players fit for competition as effectively as possible, to attend to commercial pressures.

In the 1990s, sport was a key agenda item in The Hastings Center's project on the "Prospect of Technologies Aimed at the Enhancement of Human Capacities" (Parens 1998). In this text, frequent references to sport help to elucidate what it is about enhancement that people find so alarming and which may reveal an inconsistency in moral reasoning about what sorts of things people ought to value. Discussions describe how enhancement can undermine the appreciation of human excellence or the intrinsic value of our practices, by short-circuiting the process one would typically be required to undertake when progressing from novice to expert. On this basis, scholars have described performance enhancement as cheating in at least two senses – cheating other competitors with whom a tacit contract to maintain certain rules had been made and cheating the activity, or undermining its value. Some authors even argued that sports doping contributes to human suffering by leaving our cultural pursuits more impoverished, thus extending the harm argument against the practice to broader social-psychological parameters. On such a view, sport is characterized as having value partly because it is subject to a particular kind of valued uncertainty, which is threatened by enhancement technologies since they reduce chance to bring about more predictable results.

The relationship between bioethics and sport ethics improved dramatically in the early 2000s when the genetics era provoked a series of new questions about the ends of sport and how they would interface more widely with a range of bioethical issues, focused largely on the nontherapeutic application of genetics to persons. These possibilities were reinforced by research, which suggested the possibility of characterizing performance genes, which would later give rise to specific genetic tests that claimed to identify propensity for specific capabilities in sport. In this sense, the crucial period in which bioethics and sport were brought together occurred at the beginning of the new millennium, and this period frames this entry on the relationship between bioethics and sport. These 15 years also reinforced the global nature of the problem of doping in sport and the need for sports industries to draw on bioethical investigations to achieve a more rigorous analysis of the ethics of doping. Around this time, the ethics of performance enhancement in sport became more closely aligned with wider public health concerns, where doping was seen as more than just an elite sport problem; it was also something people do to achieve a certain kind of physical ideal and identity, whether or not they played sport.

On this basis, there is value in focusing an analysis of the global bioethical issues related to

sport on this period, during which a series of key concerns were identified by a range of scholars on both sides of the ethical terrain. From here, one may advance a framework for examining bioethical issues related to sport and ensure that analyses are future focused and engaged with other key trajectories that are shaping the range of bioethical issues confronting sports. The entry proceeds by exploring the common ethical ground between sport and bioethics, before looking more closely at genetics, as a lens through which a number of global bioethics issues can be identified. Arguably, questions around genetics underpin all future ethical issues faced by sports, particularly as it concerns bioethical questions, involving the tampering of an athlete's biology through technological means and the possibility that this may have consequences for subsequent generations.

Sport in Bioethics

Over the last 15 years, bioethicists have made reference to sports within arguments about the teleology of science and medicine, particularly when inquiring into the proper role of medical practice, as a vehicle for promoting health and alleviating suffering that results from ill health, injury, or disease. For example, Chadwick (1987) uses sport as a basis for questioning the limits of health care, asking whether athletes – who take greater risks with their health than may be said of the general public – should be entitled to the same level of care as low risk-takers. In this sense, sport is discussed as an instance where one might locate limits to the role of medicine and, perhaps as a result, identify what might be the defining parameters of medical interest and responsibility. This is a problem of wider interests to bioethicists, who are confronted with expanding demands in health care.

Closely allied to this is how sport is discussed within bioethics as an example of unethical medical intervention, where, for example, genetic modification in sport would not be acceptable, since sport is too trivial an activity to require the use of such important and expensive technology, or that sporting modifications would lead to sameness and make sport uninteresting, since sport is the kind of activity that relies on creating differences. Alternatively, Ledley (1994) uses sport as an example of unethical genetic modification, arguing that its use would not satisfy Rawls' conditions of fairness, as it would further "inequalities in opportunity without a tangible prospect of benefiting those who remain at a disadvantage of furthering a state of equal basic liberties" (p. 161). Further examples of how bioethicists have drawn on sport are found in Juengst's (1998) articulation of intrinsic value in sport, which is complimented by sport philosopher Morgan's (1994) application of Alistair MacIntyre's articulations of sport ethics, where the practice of sport is defined by the distinction between internal and external goods. Where technology seems to undermine the internal goods, by collapsing the distinction between the novice and the expert, this is where one may question the value of innovation, whether it is substances, methods, equipment, or knowledge.

Another crucial moment in the drawing together of bioethics and sport is through the wider debates around transhumanism and the use of technology to make people better than well. Transhumanism is a philosophical approach that is directed toward specific kinds of technology, frequently involving discussions about emerging sciences such as cryonics, nanotechnology, or artificial intelligence. It is broadly understood as a philosophy that locates at its heart the conviction that improving the human species through technology is valuable and various proponents of transhumanism have found themselves engaged with bioethicists, debating the merit of a technologically enhanced future. Outside of sport, transhumanism has engaged bioethicists, particularly those who seek to close down conversations about the use of medical and scientific resources for anything other than nontherapeutic purposes. The application of transhumanist thought to sport has also been a point of intersection between bioethics and sport.

Other conversations that became subsumed into this posthuman future for sport included the use of laser eye surgery by golfer Tiger Woods, the development of Tommy John's surgery for

baseball players, and the wider utilization of prosthetic devices that are increasingly rivaling the athletic capacities of their biological counterparts. From 2006–2012, there was extensive debate around the Paralympian Oscar Pistorius, who campaigned to be allowed to compete in the Beijing 2008 Olympic Games, not just the Paralympic Games. In this area, bioethics contributions expanded further to include specialists in disability studies (Wolbring 2008). Such ideas were part of a wider debate about a world where technology is changing what it means to be able bodied to such an extent that one may conclude that everybody may be understood as disabled and in need of enhancement. As a bilateral below-the-knee amputee, Pistorius' 2007 campaign to compete in the Olympics – rather than just the Paralympics – was facilitated by the Cheetah legs he wears when sprinting. This innovation, and Pistorius' capacity to make it a part of his body, speaks to the way in which traditional therapeutic interventions give rise to debates about what it means to be human and how one might judge differences between people. Pistorius claimed that he is enabled rather than disabled and he should be entitled to compete alongside so-called "able-bodied" athletes. The cases also created alarm for the manner in which they suggested a future wherein healthy people might choose to amputate limbs in order to enjoy the benefits of more able prosthetics, a position which does not sit neatly with the ethics of medical intervention.

Thus, the focus on human enhancement within the bioethics and sport literature is underpinned by long-established tensions about what kinds of lives are worth living and what kinds of practices are possible to undertake in an ethically sound manner. If one takes an evolutionary perspective on our development of biotechnology, then one might consider enhancing evolution to be justified and valuable, and likely to ensure our ongoing flourishing as a species. Alternatively, if one considers that our cognitive capabilities has led to our ability to disrupt evolution in a way that compromises fundamental principles that have ensured our survivability to date, then there are crucial considerations to be made before taking enhancement any further. For bioethics, questions about the proper role of medicine, the responsibilities of the health-care industries, the importance of promoting autonomy, the distinction between therapy and enhancement, and the practice of scientific research are all issues engaged by the sport case. Indeed, sport is an appealing subject to focus on, since it often involves testing these boundaries and the limits of technology. Furthermore, for sports, the human enhancement case helpfully focuses debate on the value of human lives in practical ways, asking us to consider what matters about the cultural practices people enjoy, what becomes of a world where people are required to undergo endless medical testing to ensure fairness and justice within society, and what may be wrong with manipulating biology.

These bioethical inquiries into sport are most extensively articulated through the subject of genetic technology which, since the turn of the millennium has mobilized bioethicists and sport ethicists, each of whom have sought to understand what may be valuable or alarming about a world where humans have the capacity to modify themselves genetically.

Genetics and Sport: A Unifying Subject

In 1998, *New Scientist* magazine published an article referencing research by Montgomery et al. (1998) that suggested the existence of performance genes, which may be possible to isolate – and perhaps modify – to promote specific athletic capacities. The article came at the beginning of the public hype of the Human Genome Project, which spawned countless debates about what brave new world would ensue from an era of human genome sequencing and possible modification. By 2001, the International Olympic Committee had convened a working group looking at the possible abuse of gene therapy in sport, and in 2002, the World Anti-Doping Agency held a landmark meeting on gene doping, which was a crucial catalyst in the drawing together of expertise in sport ethics and bioethics.

One of the key determinants of this unification was the realization that athletes might use genetic

technologies to enhance their performances and, as a result, undermine the aspirations of the anti-doping industry, which, by then, had become a worldwide pseudo-political movement. However, it is also reasonable to claim that genetic scientists were equally concerned about the misuse of genetic research that this would entail and its impact on the public perception of a science that was already controversial and communicating poorly with the public. The prominence of world-renowned genetic scientist Ted Friedmann, who was named the President of the American Society of Gene Therapy in 2005 and who became a key figure as Chair of WADA's Gene Doping Expert Group, is indicative of this closer relationship between science and sport. Along with the prominent work of Lee Sweeney – also a member of WADA's group – whose research into IGF-1 would engage the world's media on its possible application to sport, the context for bringing bioethicists together with sport ethicists was ripe.

This focus was reinforced by the interventions of numerous bioethicists, whose foray into sport ethics issues – largely through discussions about genetic modification – brought greater prominence to the issues and the relationships across practice ethics. The work of Michael Sandel, Bill Joy, Ronald Green, Andy Miah, Thomas Murray, Julian Savulescu, Soren Holm, John Harris, and later Michael McNamee, Claudio Tamburrini, Gregor Wolbring, Bengt Kayser, and Alexandre Mauron, went some way to bringing sport into the frame of interest within bioethics. It is as if the genetic enhancement topic became a catalyst for, especially, controversial bioethicists or philosophers to become engaged by the applications to sport, and the policy context within the sports industry was quickly being agitated by these prospects. The consequence of these discussions was the establishment of a new community of bioethical perspectives on the enhancement debate in sport.

By 2003, WADA had included a provision within its World Anti-Doping Code to ensure that gene doping was prohibited, before anyone had any idea about whether it was even possible. At the same time, various anti-doping authorities began to invest into genetics as a basis for next-generation doping detection, while also investing funds into bioethics work to establish the foundation for prohibiting its use, mostly focused on the doping-like applications of genetics, rather than the bigger implications that might ensue from germ line engineering. Moreover, some nations were beginning to experiment with genetic technologies, hoping for them to yield more effective ways of identifying talent, but even these applications of genetics were beginning to generate controversy, and so, nearly as quickly as they came about, a culture of silence ensued about such use.

These inquiries gained further political momentum via Francis Fukuyama's varied contributions within the biopolitical sphere, along with the broader works of the US President's Council on Bioethics, the latter of which considered the sport case as part of broader debates on the prospect of human enhancements. Fukuyama's contribution to these discussions developed his End of History thesis from the mid-1980s, drawing attention to its failure to come to terms with the teleology of scientific development. Here, he argues that, because humanity has not yet seen the end game of the scientific method, one cannot fully account for the disruption to fundamental, normative structures within society that this will cause. In 2002, Fukuyama's "Our Posthuman Future" articulated this imminent demise whereby the commercialization of lifestyle biotechnological modifications to humans would compromise a proposed human dignity that, for now, holds societies together. A good example of this may be the perception that the world is predicated on some sense of the distribution of goods through merit and where enhancements could completely collapse this normative structure. In this same period, Fukuyama's work on the US President's Council on Bioethics detailed how elite sports cultures would be among the first – if not the first – physical cultures to feel the impact of this scientific revolution. In short, athletes' bodies were already marked for an imminent overhaul via biotechnology. These details function as a prologue to explaining how elite sports became the subject of extensive bioethical interest and debate, and as a result, they frame this entry's articulation of a global bioethical framework around sport.

As mentioned earlier, a number of bioethical issues have been discussed through genetic technology, which, together, elucidates a bioethical framework for sport. For example, consider the story that concerned five British footballers, who planned to store the stem cells of their children to protect themselves (and their children), should they become injured during competition. In this situation, a number of complex issues arise that reflect the cultural context within which genetic technology has developed. First, the footballers' intentions imply an acceptance of the legitimacy to harvest stem cells, which is not something that is universally shared. Second, the fathers presume an entitlement to utilize the cells of their children for their own means, rather than consider that these cells belong only to the child or, perhaps also to the mother. Third, the possibility of undertaking such a decision exists in a country where the industry of commercial stem cell storage exists, which is a possibility that is not available to all nations.

Establishing the ethical terrain of this case is undoubtedly a job that requires a bioethical framework, as there is little from the world of sports ethics that can contribute. This is not to say that sport ethicists are ill equipped to resolve such a matter, but simply that looking toward the ethical practice of sports does not shed much light on what ought to be permitted in this case. Justifying such use of genetic material on the basis of competitive advantage seems to fail to take into account the complex chain of ownership that lies behind this choice. After all, if one concludes the blood belongs to the child, then it is difficult to achieve consent to its use. Yet, there are implications for sports, should such actions be permitted. After all, if such means are considered reasonable to promote recovery from injury, then the level playing field argument within sports may require that sports involve themselves in promoting the harvesting of cord blood, so that all athletes have a better – and equal – chance of recovering at the same pace when suffering injury. This seems an incredibly unlikely scenario, and yet, the case reminds us that what takes place outside of sport has an impact on what kinds of values sports can uphold. If stem cell harvesting is widely used outside of sport and if its use can allow an athlete to transition from recovery to, perhaps, better than well performance, then it is hard to see how sports prohibit such use. Equally, if society permits germ line genetic enhancements, which have the consequence of creating enhanced offspring, then it is unlikely that the world of sport can do much to prevent this and, by implication, prohibit such people from participating in sport on the basis of some appeal to ethics or integrity.

The example also allows us to recognize that established ethical codes of conduct are not adequately constructed to accommodate a number of novel scientific applications. This need not mean that societies devise new ethical codes to take into account such beliefs, but it might require taking into account special sensitivities that arise because of the importance people give to genes. It might also imply revising ethical protocols, such as the consent process, to account for the different ways in which people make sense of genetic information. Arguably, this is why the importance of genetic counseling has been emphasized in a clinical context, as these sensitivities give rise to new obligations for the world of sport to address.

Ethics for the Genomic Era

Understanding how best to approach these questions involves looking back on how ethical theory has changed in a post-genomic era. In the context of medicine and technology, early approaches to bioethics and medical ethics focused on developing principles that could govern good practice, which dominated the rise of bioethics in the post-war period. Their four principles – autonomy, beneficence, non-maleficence, and justice – have shaped the development of ethical codes within modern medicine and science. In recent times, scholars have critiqued this top-down principlism, arguing that the lived reality of ethical conduct is much more complex and that ethical codes must be informed by these circumstances. Thus, a bottom-up approach to deriving ethical guidelines has also emerged, so-called casuistry (the study of cases). Today, a method of *reflective equilibrium* has become commonplace within applied ethical settings, which relies on a combination of the two approaches, while increasingly empirical research

is informing the characterization of ethical dilemmas within the medical setting.

Another bioethical controversy that surrounds the use or development of genetic science in sport is the distinction between research and application. For example, the aforementioned Lee Sweeney's research on insulin-like growth factor 1 (IGF-1) aspires to treat muscle-wasting diseases. Yet, Sweeney's work has also been at the forefront of the gene doping debate – much more than the research might otherwise have been – in part due to his willingness to engage publicly on how the future of his work could be utilized (Barton-Davis et al. 1998).

There are many other genetic scientists whose work holds similar implications and this presents challenges for anti-doping policy makers. This is because the kind of developmental work that goes into medical research is highly protected until it is commercialized. Yet the capacity of the sports world to address illegal uses of such technology relies on early indications of the products that are likely to emerge on the market, to ensure it can develop robust anti-doping tests. In addition, the controversy surrounding gene doping has meant that experimental research surrounding genetics and exercise science has also met with skepticism within the policy community. As a result, the British Association for Exercise and Sport Science published a position statement arguing on behalf of genetic research in sport (Wackerhage et al. 2009). The authors assert that there are novel challenges presented by genetic science in sport, but that there should be encouragement to continue. For example, one of the difficulties with separating research from application is the unexpected knowledge that might derive from research, as time goes on. For example, they discuss how

> ..a polymorphism in the gene encoding the human bradykinin receptor B2 was shown to be associated not only with exercise-induced cardiac hypertrophy (Brull et al. 2001) but later that it also predicted coronary risk (Dhamrait et al. 2003). (ibid: 1113)

Genetic Information as an Ethical Guide

Outside of sport, the use of genetic information is also taken very seriously, as its abuse may threaten the enjoyment of certain human rights. The apparent desire of employers or insurance companies to have access to genetic information, as a tool for limiting their economic risk, invites scrutiny due to its potential for discrimination. Moreover, the possibility of identifying specific genetic characteristics, coupled with the possibility of selecting in or out certain traits, may lead to considerable social pressure to undertake such decisions. Indeed, one might envisage that such choices become an integral part of what is deemed to be responsible practice in sport talent identification.

Again, these circumstances reveal the broad societal concerns that orbit the use of genetic science in sport, while also reinforcing the importance of a bioethical approach to the sports technology problem. Genetic testing in sport has arisen in two prominent cases. The first is outlined by the *Australian Law Reform Commission* (ALRC), which conducted the first major investigation into the use of genetic information in sport. Here, they outline the case of the Professional Boxing and Combat Sports Board of Victoria, which discussed whether genetic testing could be used to help physicians advise (or better inform) the athlete about the level of risk in their competition. McCrory (2001) mentions a similar concern, explaining how "delayed cerebral oedema after minor head trauma" has been linked to "an abnormality with the CACNA1A calcium channel subunit gene" (p. 142). McCrory argues that these findings are important enough to require physicians to offer advice against participation in sports and even to require athletes to take genetic tests, where such risk exists. Moreover, the ALRC note that a "milder form of this condition can occur in players of rugby, soccer, and other sports associated with repetitive blows to the head" (section 38.29, p. 964). Yet, the Boxing Board decided against the use of such tests. A second case was of Ed Curry, basketball player for the Chicago Bulls team, who was required to undertake a genetic test after he had missed games due to an irregular heartbeat. Curry refused the test and transferred to the New York Knicks in 2005, a club that did not require him to take such a test.

The two examples highlight the complexity of maintaining bioethical principles – such as

confidentiality – in what are often high-profile cases. Moreover, they highlight the legal uncertainty within the sports world over how to address claims from a range of parties over access to genetic information. Alongside these debates, a number of conversations have also taken place about the use of genetic information to make selection decisions on the basis of *athletic potential*, rather than liability of health risk. For some time, there was ambivalence about the legitimacy of such selection decisions when even the International Olympic Committee President Jacques Rogge reportedly indicated that there is nothing obviously unethical about refining talent identification techniques using genetic information.

However, the eventual realization of such testing has provoked sport leaders to revise their position on such use. This was most apparent in 2004, when the first commercial test for a performance gene reached the market, called the ACTN3 Sports Performance Test™. It claimed to identify whether the user may be naturally geared toward sprint/power events or toward endurance sporting ability. Around the same time, *Nature* reported that an Australian rugby league team would experiment with genetic tests to improve their ability to train athletes and direct them toward success within competition (Dennis 2005). Soon after, WADA responded with its *Stockholm Declaration*, which "strongly discouraged" the use of such tests by sport organizations, especially to make selection decisions.

The use of genetic information reinforces the broader societal implications of the genetics issue in sport. Such concerns can involve extending the realm of parental autonomy, though in dramatically different ways. For prenatal selection decisions, it would involve presenting parents with decisions about what kind of embryo they bring into existence. Alternatively, in a postnatal setting, it can imply using a mouth swab to identify what sports children might excel in. The ethical and moral objections to such technology being used range from concerns about engendering an endless spiral of biotechnological competitiveness to anxieties that such selection decisions exhibit unjustified and inappropriate prejudices toward certain kind of people over others. Yet, these freedoms may also fall well within the accepted freedoms of parental liberty and so do not present such great harms as to require prohibition.

Practical Sport Ethics as Bioethical Test Cases

A number of practical bioethical questions also emerge from such prospects, which have yet to be resolved. For example, would athletes be required to undergo genetic screening to establish their genetic profiles before being allowed to compete in sports? Alternatively, in what ways could sport authorities and their stakeholders have access to the information derived from genetic tests to identify doping practices? How would genetic testing influence an athlete's enjoyment of sport? If an athlete has an unusually favorable phenotype for a given sport, would this lead to their disqualification from competition on the basis of their having an unfair advantage?

Where genetic testing is used to identify talent, concerns over discrimination are of a different character. Here, the concern is that the testing method may not be an adequate indicator of performance potential. Indeed, the complexity of sports is such that making absolute judgments about what characteristics will ensure or even increase the probability of success is difficult. For example, one may reasonably argue that extreme shortness (or extreme height) prevents an individual from performing the required skills of many sports. As such, by claiming that height is a relevant characteristic of sporting performance, one may then claim that genetic tests could be used to justify why short children are not selected in an elite training program. Yet, this conclusion is insufficient, since there are many people who may welcome the chance to become an elite athlete, but for whom there is no opportunity due to the tests. Indeed, such arguments could be made in relation to a number of disabilities, for which it would be unreasonable to claim that such aspirations do not deserve support.

Thus, clearly there is a sense in which sport depends on providing opportunities for different kinds of people. Moreover, it seems preferable to adapt the structures in sport to allow the possibility of such people to pursue elite competition,

rather than to endorse a system, which excludes certain kinds of people from participating. If this additional commitment requires creating greater divisions within sports, then this should be the responsibility of sport federations, since the value of sport depends on inclusivity. An additional complication in the context of sport is taking into account the life course of athletes, where participation in sport often starts at a very young age. This has a specific bearing upon the use of such testing and selection in children since a child may enjoy many years as a competitive athlete, before reaching a point where genetic factors limit competitiveness at an adult level. Consider a young basketball player who is destined to be 165 cm tall, who may enjoy being an excellent player until his peers have undergone their final growth spurts. The value a child may accrue from these experiences is clearly sacrificed, were a genetic test to be used early in life reveal that his eventual height would likely limit competitiveness in adulthood.

If discussions about genetic information reflect the present-day use of genetic information, conversations about gene transfer in sport are its future. These debates have been dominated by the prospect of *gene doping*, the use of gene transfer for nontherapeutic or enhancing purposes. A range of institutions take this prospect seriously and include WADA, the American Association for the Advancement of Science, the US President's Council on Bioethics, and the British Government (House of Commons 2007). A number of philosophical and ethical issues surround the debate on gene transfer in sport. On a philosophical level, there is a need to distinguish between types of therapy and non-therapy. For example, while the WADA Code accepts the use of gene transfer for therapeutic use, it is unclear whether the distinction between therapy and enhancement can be sustained in the long term. Thus, insofar as genetic disorders are often linked to age-related diseases – such as muscular depletion – it might be medically desirable to "enhance" people in order to maintain a reasonable level of health. Moreover, it might appear that individuals must be treated with gene transfer well before they are symptomatic, that is, when they are considered healthy. These prospects are receiving careful consideration from a range of governments around the world. As noted in the introduction, the US President's Council on Bioethics discussed this prospect in the context of sport and identified no clear consensus on what should follow in policy terms. In addition, the British House of Commons (2007) investigation into human enhancement technologies in sport also reinforced the likely expansion of such modifications in society.

Ethically, the recurrent questions are about how such technology would affect equality in sport and broader notions of justice. WADA's approach to any new technology is to identify whether it engages two of the three of its criteria: performance enhancing, risky to health, and against the spirit of sport. If two conditions are engaged, then it will *consider* prohibition. While the use of gene transfer remains highly experimental, it may give rise to forms of performance enhancement that are safer than current methods that rely on synthetic substances – often from an illegal black market. On this basis, there might be good reason to promote these healthier modes of human enhancement in order to diminish the illegal trade of substances that currently overshadows elite sports. Moreover, through the utilization of biomarkers and DNA passports, there might be a greater potential to monitor the detrimental health risks that an athlete could face through such modifications. These arguments are part of broader perspectives that argue on behalf of human enhancement in sport, which have gained prominence in recent years.

Conclusion

Within sports, the bioethical debate about human enhancement is configured in different ways, compared with their use more generally. While many interventions still require the approval of medical authority and, thus, are still subject to medical ethics, there are crucial differences. First, the pursuit of performance enhancements is central to the logic of elite sport competitions, where the importance of winning is paramount for

many reasons. The cultural and global edifice of elite sport undoubtedly places great emphasis upon the ability to demonstrably excel. As such, opportunities to enhance performance are central to the praxis of elite competitions.

Perhaps the only limiting factor for the full pursuit of transhuman enhancements is the aspiration to ensure fairness within competition, another of sport's central values. Yet, in response, one might conclude that sport authorities should legitimize a greater number of performance modifiers, thus making them legal and silencing those who argue against such enhancements on the basis of fairness or concerns about cheating. Athletes engage in highly sophisticated forms of training and use various kinds of legal, transhuman technologies to adapt to and cope with different environmental demands, such as the physical stress of swinging a tennis racquet on the shoulder. Indeed, the Olympic motto reads as *Citius*, *Altius*, *Fortius* – Faster, Higher, Stronger – which perpetuates the transhuman values under discussion. Moreover, to the extent that it is becoming increasingly difficult to achieve greater levels of performance without new technology, this is further reason to pursue such modifications. Indeed, athletes must be allowed to discover new means of performance enhancement to take sport performances to new levels. In turn, this will require the acceptance of many technologies that are considered, outside of sport, to be ethically problematic for medical professions to accept. To help sports get to this position in a safe and ethical manner, there may be a need for something like a World Pro-Doping Agency to complement the World Anti-Doping Agency, the responsibility of which is to discover and invest into developing safer forms of athletic enhancement.

The fact that there is an apparent moral community that is opposed to certain forms of performance enhancements, such as steroids or other doping technologies, does not discount the fact that sports are awash with many forms of legal enhancement. Indeed, the technological status of sports is self-evident, with many kinds of sports equipment accepted that have transformed athletic performance. Furthermore, the degree of sophistication that sport equipment describes speaks to technology that will soon become indistinguishable from the athlete's body. Consider again what might come after the enabling technology of Oscar Pistorius' prostheses. Conceivably, future athletes will pursue elective surgical interventions to achieve other enhancements, such as invasive leg extensions. Indeed, this already takes place outside of sports. At the very least, it is probable that injured athletes would opt for enhanced prosthesis, rather than to just seek repair to a level of normal functioning. Such athletes may then return to competitive sport, putting themselves at an advantage over their biological counterparts. In this world, both bioethics and sport ethics are key to resolving not only what is just but also what is important about the means by which people achieve certain goals and the range of people who are involved in bringing them about. Undoubtedly, the future of human enhancements in sport is characterized by the greater involvement of scientists, suggesting that future winners will be attributed much more to the athlete's entourage. However, neither bioethics nor sport ethics has adequately shown that this will be a less engaging form of human competition, and one view of present-day sports is that it is already a competition among scientists and technologists. Furthermore, a world where the 100 m sprint is won in 5 s may attract even more television viewers than it does presently, and the increased popularity of prosthetically enhanced Paralympic sport may be indicative of such interests and the worth of such pursuits.

Cross-References

- ▶ Casuistry
- ▶ Doping
- ▶ Enhancement
- ▶ Principlism
- ▶ Transhumanism

References

Barton-Davis, E. R., Shoturma, D. I., Musaro, A., Rosenthal, N., & Sweeney, H. L. (1998). Viral mediated expression of insulin-like growth factor I blocks

the aging-related loss of skeletal muscle function. *Proceedings of the National Academy of Sciences U S A, 95*, 15603–15607.
Chadwick, R. F. (1987). *Ethics, reproduction and genetic control*. London: Routledge.
Coenen, C., Schuijff, M., Smits, M., Klassen, P., Hennen, L., Rader, M., & Wolbring, G. (2009). Human Enhancement: Study. Brussels: Science and Technology Options Assessment, European Parliament.
Dennis, C. (2005). Rugby team converts to give gene tests a try. *Nature, 34*, 260.
European Parliament: Science and Technology Options Assessment. See: http://www.europarl.europa.eu/stoa/events/workshop/20090224/background_en.pdf
Fost, N. (1986). Banning drugs in sports: A skeptical view. *Hastings Center Report, 16*, 5–10.
Fukuyama, F. (2002). *Our posthuman future: Consequences of the biotechnology revolution*. London: Profile Books.
House of Commons. (2007). Human Enhancement Technologies in Sport: Second Report of Session 2006–7 (22 February). (Science and Technology Select Committee, Ed.). The Stationary Office, London.
Juengst, E. T. (1998). What does enhancement mean? In E. Parens (Ed.), *Enhancing human traits. Ethical and social implications* (pp. 29–47). Washington, DC: Georgetown University Press.
Ledley, F. D. (1994). Distinguishing genetics and eugenics on the basis of fairness. *Journal of Medical Ethics, 20*, 157–164.
McCrory, P. (2001). Ethics, molecular biology, and sports medicine. *British Journal of Sports Medicine, 35*, 142–143.
Montgomery, H., Marshall, R., Hemingway, H., Myerson, S., Clarkson, P., Dollery, C., & Humphries, S. E. (1998). Human gene for physical performance. *Nature, 393*, 221–222.
Morgan, W. J. (1994). In B. G. Rader & R. Roberts (Eds.), *Leftist theories of sport: A critique and reconstruction*. Urbana: University of Illinois Press.
Murray, T. H. (1983). The coercive power of drugs in sports. *Hastings Center Report, 13*, 24–30.
Parens, E. (1998). *Enhancing human traits: Ethical and social implications*. Washington, DC: Georgetown University Press.
Wackerhage, H., Miah, A., Harris, R. C., Montgomery, H. E., & Williams, A. G. (2009). Position stand on 'Genetic research and testing in sport and exercise science': A review of the issues. *Journal of Sports Science, 31*, 1–8.
Wolbring, G. (2008). Oscar Pistorius and the future nature of Olympic, Paralympic and other sports. *ScriptEd, 5*(1), 139–160.

Further Readings

McNamee, M. (2007). Whose prometheus? Transhumanism, biotechnology and the moral topography of sports medicine. *Sport, Ethics and Philosophy, 1*(2), 181–194. doi:10.1080/17511320701425173.
Miah, A. (2003). Be very afraid: Cyborg athletes, transhuman ideals, and posthumanity. *Journal of Evolution and Technology, 13*(2). http://www.jetpress.org/volume13/miah.html. Last Accessed 22 May 2015.
Miah, A. (2004). *Genetically modified athletes: Biomedical ethics, gene doping and sport*. London/New York: Routledge.
Schneider, A. J., & Friedmann, T. (2006). *Gene doping in sports: The science and ethics of genetically modified athletes*. New York: Academic.

Standards of Care

Keymanthri Moodley
Department of Medicine, Centre for Medical Ethics and Law, Stellenbosch University, Cape Town, South Africa

Abstract

Established treatment and/or prevention interventions exist for most medical disorders. These may be single interventions or they may comprise a constellation of health interventions and are widely regarded as "standards of care." These standards range from no treatment (especially in resource-poor settings) to a gold standard that is international, expensive, and complex. In the context of international multisite research conducted by resource-rich countries in resource-poor countries, standards of care used in the control arm of the study are often controversial especially when placebo is used in this group of participants. Such controversy resulted in global debate in the 1990s when antiretroviral treatment for pregnant women was tested against placebo in several resource-poor countries despite the establishment of a gold standard of care in resource-rich countries. Charges of ethical imperialism, ethical relativity, and exploitation of vulnerable populations were expressed in global debates.

Similar arguments emerged in the context of surfactant trials in premature infants in Bolivia. Guidance regarding standards of care in control groups enshrined in the Declaration of Helsinki became extremely controversial in the context of these global debates and increased the sensitivity

of research ethics committees to the standard of care being used in clinical trials generally.

This chapter discusses the evolution of the debate on standards of care in clinical trials and adds to the controversy that abounds in research ethics.

Keywords

Research; Standards of care; Placebo; Control arm

Introduction

There are established treatment and prevention strategies in clinical health care for most medical disorders that are regarded as "standards of care." These standards range from no treatment (especially in resource-poor settings) to a gold standard that is usually international, expensive, and complex. Such standards differ from country to country. They also vary within countries from private to public health systems. As new scientific evidence emerges from medical research, such standards of care may change. The ethical debate around standards of care in research was advanced in the context of placebo-controlled randomized controlled clinical trials conducted in developing countries to prevent HIV transmission from pregnant women to their babies and in the context of surfactant trials in premature infants in Bolivia (Lindsey et al. 2013).

History and Development

In 1994, the results of the first randomized placebo-controlled study on pregnant women infected with HIV were published. It was established that intensive treatment of these women with the antiretroviral drug zidovudine during pregnancy and delivery reduced the transmission of the virus from mother to child by 67 %. From this point onward, zidovudine became the best proved standard of treatment for all HIV-infected pregnant women in the United States (Connor et al. 1994).

The drug regimen used in this landmark study was, however, very expensive and unaffordable to Third World countries. The next logical step was therefore to investigate the possibility of shorter and hence cheaper courses of treatment. The World Health Organization (WHO) urgently called for research in developing countries to explore simpler and less expensive drug regimens. The Joint United Nations Programme on HIV/AIDS (UNAIDS) and other organizations collaborated to set up 16 clinical trials in 12 developing countries around the world. Nine of these studies were conducted by the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC). One of these trials (conducted in Thailand) was designed as an equivalence study – three short-course regimens were compared and the control group was given the ACTG 076 regimen. However, 15 of these 16 trials were randomized and placebo controlled. HIV-infected pregnant women in the study group were given a short course of zidovudine, and the incidence of transmission of the virus to their babies was established. However, the HIV-infected pregnant women in the control group were given a placebo, which was where the controversy began (Lurie and Wolfe 1997).

The Placebo Debate

In April 1997, Dr. Peter Lurie and Dr. Sidney Wolfe of the Health Research Group (an arm of the watchdog organization, Public Citizen) sent a letter to the secretary of the Department of Health and Human Services, Donna Shalala, which stated the following:

> Unless you act now, as many as 1002 newborn infants in Africa, Asia and the Caribbean will die from unnecessary HIV infections they will contract from their HIV-infected mothers in nine unethical research experiments funded by your department through either [NIH or CDC].

In September 1997, Lurie and Wolfe repeated their charges in the *New England Journal of Medicine*. They drew attention to the two studies being conducted in the United States where patients in all study groups had unrestricted access to

antiretroviral drugs unlike the 15 short-course trials in developing countries where women in the control group were given a placebo (Lurie and Wolfe 1997). The editorial in the same issue of the journal written by the executive editor, Dr. Marcia Angell, supported the views of Lurie and Wolfe. In addition, she drew a parallel between withholding treatment in the placebo group and withholding treatment for syphilis in the infamous Tuskegee study. This set in motion an unprecedented debate on the vertical transmission trials and the ethics of collaborative multinational research (Angell 1997).

The Scientific Debate

The Research Question and Clinical Equipoise

Lurie and Wolfe argued that by conducting a placebo-controlled trial, the researchers were, by implication, asking the wrong question:

> Is the shorter regimen better than nothing?

The presumed answer to this question was that anything would be better than nothing. It is an essential prerequisite that when a randomized clinical trial compares two different treatments for a disease, there should be no good reason for thinking that one is better than the other. Hence, investigators need to be in this state of clinical "equipoise" when embarking on a randomized clinical trial. If there is any evidence that one option might be better than the other, then:

> not only would the trial be scientifically redundant, but the investigators would be guilty of knowingly giving inferior treatment to some participants in the trial. (Angell 1997)

Hence, randomized clinical trials create the potential for conflict between the investigator's role as doctor and research scientist. During recruitment, a doctor must ask a patient to submit himself or herself to random assignment to one of two different treatments, one of which may be a placebo. This request can only be ethically justified if the researcher is in a state of genuine uncertainty regarding which treatment is better. This is so because randomization is inconsistent with doing one's best for the patient as a doctor (Miller and Weijer 2003). This rule applies to placebo-controlled trials in that it is only ethical to compare a potential new treatment with a placebo when there is no known effective treatment.

In the opinion of Lurie and Wolfe, the question that should have been asked was:

> Can we reduce the duration of prophylactic [zidovudine] treatment without increasing the risk of perinatal transmission of HIV, that is, without compromising the demonstrated efficacy of the standard ACTG 076 [zidovudine] regimen? (Lurie and Wolfe 1997)

In response to this charge, Varmus and Satcher retorted that they were looking to answer a much more complex question than Lurie and Wolfe suggested. Their concern was not simply to establish whether a short course of treatment was better than nothing but also whether the short course was safe and if so whether the demonstrated efficacy compared to placebo was large enough to make it affordable to the governments in question. This viewpoint was supported by an internationally renowned South African HIV researcher who argued that the fundamental research question related to whether short courses of antiretrovirals could reduce vertical transmission sufficiently to warrant their wide-scale implementation in South Africa (Abdool Karim 1998).

Varmus and Satcher argued further that:

> the most compelling reason to use a placebo-controlled study is that it provides definitive answers to questions about the safety and value of an intervention in the setting where the study is performed, and these answers are the point of the research. (Varmus and Satcher 1997)

The investigators believed that two different populations were being studied, and it was not possible to extrapolate findings from the United States to Africa. The ACTG 076 regimen in the United States required that women receive HIV testing and counseling early in pregnancy, comply with oral treatment for several weeks and intravenous antiretrovirals during labor, and refrain from breast-feeding. In addition, babies would have to receive six weeks of oral antiretrovirals. South Africa, in common with other developing countries, had a high frequency of home deliveries especially in rural communities (Abdool Karim

1998). In developing country settings, women present late for antenatal care have limited access to HIV testing and counseling and depend on breast-feeding to protect their babies from malnutrition and diarrheal diseases. The safety of zidovudine in populations who have a high incidence of malnutrition and anemia was unknown. The cost of the ACTG 076 regimen was approximately $800 per treatment, far in excess of the per capita health-care expenditure of under $10 in most developing countries (Varmus and Satcher 1997). Charges were also made that the critics' commentary of the trials "reflects a lack of understanding of the realities of health care in developing countries" (Halsey et al. 1997).

The Utility of Existing Data

There was disagreement on the use of observational or historical data to provide the same information that could be obtained from the placebo arm. Advocates of placebo-controlled trials and the WHO argued that "historical controls" were not reliable sources of data due to the change in vertical transmission rates from one country to another. Abdool Karim agreed and substantiated his claims with data from South Africa that indicated differences in vertical transmission rates from 1991 to 1994. He added that the vertical transmission rate is influenced by a number of factors including cesarean section rates, maternal viral load, and breast-feeding rates. As such, the use of historical controls would lead to spurious and hence unacceptable conclusions (Abdool Karim 1998).

Critics of the trials however believed that the differences between the ACTG 076 trial participants and those in sub-Saharan Africa were being exaggerated and that HIV vertical transmission rates were known in Africa and were in the region of 20–30 %, making the use of historical controls possible.

Equivalence Trial Issues

When effective treatment exists, a placebo may not be used, and subjects in the control group must be given the best known treatment (Angell 1997). Such a study is termed an equivalence study and the results are scientifically valid.

If the ACTG 076 regimen were used as the control group in the controversial vertical transmission trials, it would be termed an equivalence trial. Such a trial would be useful if it were proven that short-course treatment regimens were as good as or better than the ACTG 076 regimen. It is also necessary for the expected outcome of the control to be known. Abdool Karim argued that the effect of ACTG 076 in South Africa is not known and could not be extrapolated from other settings given the differences in breast-feeding rates, sexually transmitted disease rates, cesarean section rates, levels of viral load, and other variables.

Advocates of placebo-controlled trials held that equivalence trials required a much larger sample size to show a difference between two active arms of the study, and hence, they would take longer to complete and cost more. Furthermore, the larger numbers of participants would result in exposure of more people to the risk of research.

Ethical Dimension

The Guidelines

Critics of placebo-controlled trials argued that the trials violated principles enunciated in several major international ethics guidelines. The Declaration of Helsinki was exhaustively invoked. In support of her objections to the placebo-controlled trials, Angell cited the following tenets of the DOH 1996:

> In research on man, the interest of science and society should never take precedence over considerations related to the wellbeing of the subject.

and:

> In any medical study, every patient – including those of a control group, if any – should be assured of the best proven diagnostic and therapeutic method. (WMA 1996)

Guidelines 8 and 15 of the WHO document – CIOMS (1993) – were frequently invoked.

Here, researchers were required to ensure, inter alia:

> that persons in underdeveloped communities will not ordinarily be involved in research that could be

> carried out reasonably well in developed communities and that research was responsive to the health needs and priorities of the community in which it is to be carried out.

Guideline 15 stated that the proposed study should be submitted for ethical and scientific review, and the ethical standards applied "should be no less exacting than they would be" for research in the sponsoring country itself (CIOMS 1993).

Advocates of the placebo trials cited the principles of the Belmont Report. Emphasis was placed on the shift from the principle of beneficence to justice – equitable access to clinical trials (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Guideline 8 of the CIOMS document relating to responsiveness to local needs in research conducted in developing nations was also cited. Hence the guidelines were used as ammunition to defend the positions of both proponents and critics of the placebo-controlled trials indicating the internal contradiction that exists in many international documents.

Standard of Care

After the efficacy of the ACTG 076 regimen had been established in the United States in 1994, it became the "gold standard" in the prevention of mother-to-child transmission of HIV. Hence, both critics and proponents of placebo-controlled trials were in agreement that placebo-controlled trials could not be conducted in the United States. Critics of the trials argued for a universal standard of care irrespective of where in the world the research was being conducted.

Proponents however argued that participants in the control group would have received exactly the same standard of care if they had not participated in the trials – the local standard of care which at that time was no treatment in the developing world.

According to Marcia Angell, the justifications for these trials are:

> reminiscent of those for the Tuskegee study: Women in the Third World would not receive antiretroviral treatment anyway, so the investigators are simply observing what would happen to the subjects' infants if there were no study. And a placebo-controlled study is the fastest, most efficient way to obtain unambiguous information that will be of greatest value in the Third World. (Angell 1997)

Ethical Imperialism

Ethical universality refers to the belief that the ethical principles that guide the conduct of research are the same wherever in the world research is conducted. Ethical relativism refers to the belief that ethical principles that guide the conduct of research vary from one cultural setting to another. This concept is based on skepticism and tolerance. Skepticism refers to the belief that actions may be defined as right or wrong by specific people in specific cultural contexts at specific times. Hence behavior is culturally relative. Ethical relativity contends that the "impossibility of objectively determining moral action obliges tolerance toward other cultures" (Christakis 1996).

Hence in transcultural research, the ethical requirements of both cultures involved will need to be met. This approach is problematic in that a third cultural system could regard the two systems involved as unethical and there is no provision made for conflict resolution. Ethical pluralism on the other hand "acknowledges the key position of culture in shaping both the content and the form of ethical rules and it includes a mechanism of dispute resolution through mutual evaluation and negotiation" (Christakis 1996).

Critics of the placebo-controlled trials were accused of ethical imperialism – trying to impose their ethical standards on countries that had made their own judgments on the trials, based on their particular needs.

This debate predated the actual conduct of HIV vertical transmission trials. Marcia Angell, in 1988, raised the fundamental question of whether:

> ethical standards are relative, to be weighed against competing claims and modified accordingly, or whether like scientific standards, they are absolute. (Angell 1988)

She argued then, as she did in 1997, that fundamental principles of humane research should not be compromised. She maintained that:

> Subjects in any part of the world should be protected by an irreducible set of ethical standards, including the requirements that they not be subjected to unreasonable risks and that they be asked for informed consent to participate. (Angell 1988)

Local investigators, however, thought otherwise. Commentary from the Uganda Cancer Institute was as follows:

> These are Ugandan studies conducted by Ugandan investigators on Ugandans.

The studies in Uganda had been approved by local ethics committees.

Dr. Nicolas Meda, an epidemiologist from Burkina Faso, argued that health research in poor countries should be designed and conducted pragmatically, in keeping with local health needs and priorities. In 2002, he addressed a conference of European medical ethicists and made the following statement:

> Dogmatic interpretation of universal ethical principles in medical research will paralyse research efforts to improve HIV/AIDS prevention and treatment in sub-Saharan Africa. (Richards 2002)

Marcia Angell argued that:

> ethical imperialism obscured a more insidious danger to developing countries: ethical relativism, which opened the door to exploitation of the vulnerable peoples of the Third World.

Critics of the trials dismissed the charge of ethical imperialism and drew attention to the conflict of interest many investigators were in due to the substantial amount of research money at stake. Marcia Angell argued that researchers who levied charges of ethical imperialism against her were not necessarily advocates of the poor in their countries. Professor Hoosen Coovadia, one of the investigators involved in the Petra trials in South Africa, responded to her charge as follows:

> In these debates it was implied that we are merely passive recipients of research plans devised in Europe or the USA. This is not so, and in many instances we actively seek assistance to pursue research ideas of importance to our people. Indeed, in South Africa the barren years of apartheid isolation have instilled in us a keen appreciation of international co-operation – the HIV projects are as much ours as they are the property of our international partners. We have demanding Ethics Committees in our Universities (the first was established at the University of Witwatersrand in 1966) and regularly updated guidelines on Ethics for Medical Research published by the Medical Research Council. Our research is therefore conducted in an environment where the protection of the individual and communities is safeguarded. The assertions by Angell, Lurie and Wolfe accordingly challenge our sovereignty in making and implementing our own decisions. (Coovadia and Rollins 1999)

The debate on ethical relativism versus ethical universalism was highlighted by the attempt to apply international declarations in various developing world settings in the context of HIV vertical transmission trials.

Justice

Grodin and Annas based their objections to the trials on the principle of justice. They argued that poor participants should not bear the burdens of research that they were not going to benefit from. It was clear at the time the trials were conducted that the Ministry of Health in South Africa was not going to sanction the provision of short-course antiretroviral treatment to pregnant women even if the trials did prove the treatment to be efficacious. His argument was underscored in Minister Zuma's decision in 1998 not to provide the four-week course of treatment to pregnant women (Knox 1998). In retrospect, that was probably a good decision. After all, the four-week treatment regimen did not prove to be efficacious. However, the basic tenet of the argument remains valid – a protocol should contain a plan to implement results.

Vertical Transmission Trials in South Africa: The Results

Many of the arguments posed by both critics and advocates of the placebo-controlled vertical transmission trials were validated or rejected by the results of the trials in developing countries. I will focus my discussion on the results of the Petra trials that were conducted in South Africa, Uganda, and Tanzania between June 1996 and January 2000.

1457 HIV-positive pregnant women were randomized to one of four groups: A, B, C, or placebo. Groups A, B, and C had different short-course antiretroviral drug regimens. To facilitate

Standards of Care, Table 1 HIV transmission at week 6 and month 18

Trials	HIV transmission Week 6 (%)	HIV transmission Month 18 (%)
Petra A	5.7	15
Petra B	8.9	18
Petra C	14.2	20
Placebo	15.3	22

ease of understanding of results, the HIV transmission rates in the various groups are presented in Table 1 at week six and month 18.

The results indicate that although regimens A and B were effective in reducing HIV transmission compared to placebo, this effect could not be sustained to 18 months. This can be attributed to the predominance of breast-feeding in these populations compared to the ACTG 076 regimen study population. The investigators in this study have justified the use of placebo on the basis of a difference in study populations. They also indicate that if a placebo group had not been used or if the ACTG 076 regimen was used instead of placebo, two errors in interpretation would have occurred. The Petra C regimen would have been considered to be effective and the degree of effectiveness of all three groups would have been overestimated (PetraStudyTeam 2002). Similar results were established in the HIVNET 012 study in Uganda where breast-feeding impacted on HIV transmission rates at 20 months but to a lesser extent (Guay et al. 1999).

ACTG 076 and the Vertical Transmission Trials

How do these results correlate with the criticism leveled against these trials in 1997?

The charge of lack of clinical equipoise cannot be substantiated. If there were no clinical equipoise, the short-course treatment would have been more effective than placebo – this was only the case for 6 months after the study was initiated. Follow-up to 18 months, however, revealed no statistically significant difference between treatment and placebo groups.

The reasons forwarded regarding the differences between the North American study population and the African study populations and uncertainty regarding how they would impact on the results were borne out in this study: breast-feeding alone made an enormous difference to efficacy of short-course regimens.

The criticism regarding the use of placebo was also unjustified: there are two good scientific and statistical reasons why an equivalence trial would not have been feasible as discussed above. It is most likely that an equivalence trial would have shown that ACTG 076 was better than short-course treatment. How would that have helped the HIV epidemic in South Africa? (Coovadia and Rollins 1999).

The principle of justice was not at issue – the Department of Health did not implement the short-course treatment in 1998 – this has proved to be a good decision. The efficacy of the short-course regimen compared to placebo was not large enough to make it affordable to the South African government. When nevirapine was shown to be effective in the HIVNET 012 trials in Uganda at a fraction of the cost of other short-course regimens, treatment for pregnant women was made available in South Africa (Guay et al. 1999).

The charge leveled against critics of the placebo trials was that they were ill informed regarding health-care and research priorities in developing countries, and this appeared to be the case as was reflected in the outcome of these studies.

Finally, it appears as if the charges of ethical imperialism leveled at Angell, Lurie, and Wolfe by developing world researchers were justified, and the results of the study serve to prove that.

Tuskegee Revisited

In the course of the debates surrounding the HIV vertical transmission trials, the justification presented for use of a placebo arm was the fact that the women on the placebo arm would have received no treatment (which was the standard of care in developing countries) in the absence of the clinical trial. Marcia Angell, in her critique of the use of a placebo arm in the trials, drew the following comparison to the Tuskegee Syphilis Study:

> The justifications are reminiscent of those for the Tuskegee study: Women in the Third World would not receive antiretroviral treatment anyway, so the investigators are simply observing what would happen to the subjects' infants if there were no study.

This comparison has been challenged by the investigators involved as well as by Fairchild and Bayer. Fairchild outlines the three features in the Tuskegee study that characterize the consistent research abuses that occurred:

> First, the study involved deceptions regarding the very existence and nature of the inquiry into which individuals were lured. As such it deprived those seeking care of the right to choose whether or not to serve as research subjects. Second, it entailed an exploitation of social vulnerability to recruit and retain research subjects. Third, Tuskegee researchers made a willful effort to deprive subjects of access to appropriate and available medical care as a way of furthering the study's goals. (Fairchild and Bayer 1999)

She objects to the analogy drawn in the context of the vertical transmission trials as "investigators clearly made efforts to inform the enrolled women that they would be part of a study to reduce maternal transmission" and that some would receive placebo.

The nature of consent obtained from study participants had however been challenged by researchers working in Thailand and South Africa.

In 1998, attention was drawn to the informed consent documents used in Thailand. Discrepancies were noted in the Thai and English versions of the documents. The Thai version described the placebo as a "comparison drug that does not contain zidovudine," while the English version described the placebo as an "inactive substance" which was "like a sugar pill." The Thai critics charged that the words "inactive substance," "placebo," and "sugar pill" did not appear in the Thai documents even though Thai words or concepts did exist for these words (Achrekar and Gupta 1998, pp. 1331–1332).

In South Africa, contention was also raised by the use of the word "chuff-chuff" drug which means "pretend drug" and "spaza" drug which alludes to "half the real thing" in colloquial terms. While "chuff-chuff" drug is acceptable, "spaza" drug is misleading (Prabhakaran 1997).

Even though these controversies did exist regarding the content of informed consent documents used in the HIV trials, an informed consent process was followed in all the trials conducted in developing countries, some better than others. In no way did the HIV trials bear any resemblance to the Tuskegee study where there was an absence of the informed consent process altogether.

Fairchild goes on to contend that the social vulnerability of the women involved was not exploited. On this claim I will argue that these were vulnerable women. The UNAIDS definition of vulnerable communities includes communities with:

- Limited economic development
- Inadequate human rights protection and discrimination based on health status
- Inadequate understanding of scientific research
- Limited health-care and treatment options
- Limited ability to provide individual informed consent

The black women enrolled in the trials in South Africa definitely shared a social and economic vulnerability with the African-American men in the Tuskegee study. To the extent that this study would not have been approved in the United States on American women, an exploitation of their vulnerability cannot be denied.

However, the placebo group served only as a comparison arm for the short course, potentially more affordable regimen being tested. Tuskegee was an observational study where all participants were deprived of affordable treatment. In the HIV trials women in the placebo group were deprived of treatment that was locally both unavailable and unaffordable. In this respect, an analogy with Tuskegee cannot be drawn. Furthermore, Benatar argues that the analogy:

> minimises the deception, maleficence, paternalism, lack of accountability, racism and gross exploitation demonstrated by the researchers in the Tuskegee study. The analogy serves to trivialize Tuskegee. (Benatar 1998)

Revision of Guidance

Declaration of Helsinki 2000

As a result of the international concern evoked by the placebo debate, an attempt was made to amend the 1996 version of the Declaration of Helsinki. A proposal was made to change the specification

on treatment for control groups in the 1996 version from:

> In any medical study, every patient – including those of a control group, if any – should be assured of the best proven diagnostic and therapeutic method.

to:

> In any biomedical research protocol, every patient-subject, including those of a control group, if any, should be assured that he or she will not be denied access to the best proven diagnostic, prophylactic or therapeutic method that would otherwise be available to him or her. . . .When the outcomes are neither death nor disability, placebo or other no-treatment controls may be justified on the basis of their efficiency.

This revision was open for comment and a second debate ensued. Those who objected to the change feared that the changes would weaken the principles of the declaration:

> these revisions may inappropriately cause a shift to an efficiency-based standard for research involving human subjects and weaken the principles of the investigator's moral commitment to the research subject and the just allocation of the benefits and burdens of research, which have heretofore been the hallmarks of ethical research. The revisions will also logically lead to an explosion of research in developing countries that would be intended mainly to benefit developed countries – another affront to current notions of ethical research. (Brennan 1999)

The change to "best available" could not be implemented in the face of the strong criticism leveled against the World Medical Association. Ultimately, the change to "best current" treatment for the control group was implemented in the 2000 version.

CIOMS 2002

While the 1993 version did not include a guideline on standard of care, the 2002 version added this consideration in Guideline 11:

> As a general rule, research subjects in the control group of a trial of a diagnostic, therapeutic or preventive intervention should receive an established effective intervention. In some circumstances, it may be ethically acceptable to use an alternative comparator, such as placebo or no treatment.

There is no elaboration on "established effective" intervention – is this established globally or locally in the developing country?

Conclusion

While these major revisions were undertaken by the World Medical Association and the World Health Organization in response to the placebo debate, commentators started to question the basis for making such sweeping changes. It was charged that "tough cases make bad law," so was it valid to generalize from the placebo trials? After all the HIV vertical transmission case study had unique features – these were trials on pregnant women where placebo use meant passively allowing transmission to infants. In many cases the risk calculations in using placebo were doubled by this situation alone. It was argued that using this case study as a precedent to make revisions in guidelines that affect all research would not be valid (Brennan 1999).

The validity of this comment has been borne out in the numerous footnotes that have been added to the Declaration of Helsinki since 2000 to avoid generalization and ultimately to encourage case-by-case decisions on the use of placebo.

The revisions of both these international documents providing guidance in human participant protection evoked unprecedented attention in research ethics circles, among REC members and investigators alike. This occurred in developed and developing countries alike. Today all RECs are sensitive to standards of care used in control groups in randomized controlled clinical trials. Placebo-controlled trials are approved only where adequately justified and indicated.

Acknowledgement Material within this entry was originally part of the author's DPhil dissertation published in 2004.

Cross-References

▶ Committees: Research Ethics Committees
▶ Research: Clinical
▶ Research: Human Subjects

References

Abdool Karim, S. S. (1998). Placebo controls in HIV perinatal transmission trials: A South African's viewpoint. *American Journal of Public Health, 88*(4), 564–566.

Achrekar, A., & Gupta, R. (1998). Informed consent for a clinical trial in Thailand. *New England Journal of Medicine, 339*, 1331–1332.

Angell, M. (1988). Ethical imperialism? Ethics in international collaborative clinical research. *New England Journal of Medicine, 319*(16), 1081–1083.

Angell, M. (1997). The ethics of clinical research in the third world. *New England Journal of Medicine, 337*, 847–849.

Christakis, N. A. (1996). The distinction between ethical pluralism and ethical relativism: Implications for the conduct of transcultural clinical research. In H. Y. Vanderpool (Ed.), *The ethics of research involving human subjects* (pp. 261–278). Frederick: University Publishing Group.

CIOMS. (1993) International ethical guidelines for biomedical research involving human subjects. Geneva, Switzerland: Council of International Organisations of Medical Sciences (CIOMS).

CIOMS. (2002) International ethical guidelines for biomedical research involving human subjects. Geneva, Switzerland: Council of International Organisations of Medical Sciences (CIOMS).

Connor, E. M., Sperling, R. S., Gelber, R., Kiselev, P., Scott, G., et al. (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. *New England Journal of Medicine, 331*, 1173–1180.

Coovadia, H. M., & Rollins, N. C. (1999). Current controversies in the perinatal transmission of HIV in developing countries. *Seminars in Neonatology, 4*, 193–200.

Fairchild, A. L., & Bayer, R. (1999). Uses and abuses of Tuskegee. *Science, 284*, 919–921.

Guay, L. A., Musoke, P., Fleming, T., et al. (1999). Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of Hiv-1 in Kampala, Uganda:Hiv-1net 012 randomised trial. *The Lancet, 354*, 795–802.

Halsey, N. A., Sommer, A., Henderson, D. A., & Black, R. E. (1997). Ethics and international research. *BMJ, 315*, 965–966.

Knox, R. (1998). Despite epidemic, South Africa cuts AZT project. *Boston Globe 1*: A17.

Lindsey, J. C., Shah, S. K., Siberry, G. K., Jean-Phillipe, P., & Levin, M. J. (2013). Ethical tradeoffs in trial design: Case study of an HPV vaccine trial in HIV-infected adolescent girls in lower income settings. *Developing World Bioethics, 13*(2), 95–104.

Lurie, P., & Wolfe, S. M. (1997). Unethical trials of interventions to reduce perinatal transmission of the human immunodeficiency virus in developing countries. *New England Journal of Medicine, 337*(12), 853–855.

Miller, P. B., & Weijer, C. (2003). Rehabilitating equipoise. *Kennedy Institute of Ethics Journal, 13*(2), 93–118.

Petra Study Team. (2002). Efficacy of three short-course regimens of zidovudine and lamivudine in preventing early and late transmission of HIV-1 from Mother to Child in Tanzania, South Africa, and Uganda (Petra Study): A Randomised, Double-Blind, Placebo-Controlled Trial. *Lancet, 359*, 1178–86.

Richards, T. (2002). Developed countries should not impose ethics on other countries. *BMJ, 325*, 796.

The National Commission for the protection of human subjects of Biomedical and Behavioural Research. (1979) The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research, OPRR Reports.

Varmus, H., & Satcher, D. (1997). Ethical complexities of conducting research in developing countries. *NEJM, 337*, 1003–1006.

World Medical Association. (1996). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. Geneva, Switzerland: World Medical Association.

World Medical Association. (2000). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. Geneva, Switzerland: World Medical Association.

Further Readings

Benatar, S. R. (1998). Global disparities in health and human rights: A critical commentary. *American Journal of Public Health, 88*, 295–300.

Brennan, T. A. (1999). Proposed revisions to the declaration of Helsinki – Will they weaken the ethical principles. *NEJM, 341*, 527–531.

Prabhakaran, S. (1997). Mothers give support to placebo trials. *Mail and Guardian*: Page 5.

World Medical Association. (2000). *Declaration of Helsinki: Ethical principles for medical research involving human subjects*. Geneva: World Medical Association.

Stem Cells: Adult

Karori Mbugua
Department of Philosophy and Religious Studies, University of Nairobi, Nairobi, Kenya

Abstract

This entry focuses on some of the core ethical issues arising from human adult stem cell research. It begins with a brief exploration of the history and science behind adult stem cell research. It then addresses three main ethical

concerns about adult stem cell use, namely, the question of whether research on human embryonic stem cells should be discontinued since an ethically less controversial alternative already exists in the form of adult stem cells, the possibility of creating gametes from adult stem cells which can be used for reproduction, and the issue of stem cell tourism, a situation where desperate patients travel abroad in search of unproven stem cell therapies. Regulatory and policy issues surrounding stem cell research are also discussed. The entry concludes by saying that whereas strict ethical regulation of stem cell research is required in order to protect the safety of research participants, this must be balanced against the need to advance stem cell biology and medicine.

Keywords

Adult stem cells; Embryonic stem cells; Ethics; Stem cell tourism; Safety

Introduction

The derivation of human adult stem cells has been heralded as one of the major breakthroughs in the history of biomedicine. Adult stem cell research promises vast improvements in health care and for the first time genuinely presents the possibility of finding treatments for some of mankind's most debilitating ailments. Although adult stem cell research avoids some of the ethical issues raised by embryonic stem cell research, the proliferation in recent years of unregulated stem cell clinics around the world offering unproven stem cell interventions (both adult and embryonic) raise significant ethical concerns.

History and Biology

The term "stem cell" (in German *Stammzelle*) was coined in 1868 by Ernst Haeckel, who used it to describe the first single cells from which all multicellular organisms supposedly evolved. He also used the term to refer to the fertilized egg because it gives rise to all the cells that make up an organism (Ramalho-Santos and Willenbring 2007). However, today the term "stem cell" is used to refer to special types of cells that have the ability to divide indefinitely and to give rise to specialized cells given the right conditions. These cells can be derived from mature adult tissues and preimplantation embryos. They can also be created by reprogramming somatic cells so that they behave like embryonic stem cells. Stem cells that are created in this manner are referred to as induced pluripotent stem cells (iPSCs).

The beginnings of adult stem cell research can be traced back to 1961 when Till and McCulloch discovered stem cells in the bone marrow of mice (Till and McCulloch 1961). Adult stem cells, also known as somatic stem cells, lie dormant in most adult tissues in a microenvironment known as the stem cell niche. Their primary function in the human body is to maintain the homeostasis of tissues and to repair them when damaged by replacing the cells that have either died or lost function. Adult stem cells can be derived from infants, children, and adults as well as from the placenta and the umbilical cord after the birth of a child. In humans, these cells are also found in the bone marrow, the skin, and the intestine, among other places.

Adult Stem Cells and Embryonic Stem Cells

Adult stem cells are often contrasted with embryonic stem cells. The latter are derived from the inner cell mass of a preimplantation embryo which has to be killed in the process. Embryonic stem cells were first isolated in humans in 1998 (Thompson et al. 2008). Whereas embryonic stem cells are said to be pluripotent because they can differentiate into any type of tissue, adult stem cells are said to be multipotent, meaning that their potential to differentiate into different cell types is limited (they are lineage restricted). However, there is growing evidence that adult stem cells have greater plasticity than previously thought. For example, adult bone marrow stem cells have been found to differentiate into several

tissue-forming cells including bone, muscle, tendon, liver, kidney, and heart cells (Groove et al. 2004).

Therapeutic Potential of Adult Stem Cells

The proliferative and development potential of human adult stem cells means that they can be used to treat many degenerative diseases including Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, diabetes, and Duchenne muscular dystrophy. These diseases result from the death or dysfunction of one or a few cell types. Replacing such cells by transplantation could therefore offer long-term treatment for these ailments and greatly help alleviate the suffering of patients and their families. Other potential beneficiaries of stem cell therapy include accident victims and war casualties.

Adult bone marrow-derived stem cells have been used routinely for over 50 years to treat leukemia and other blood disorders. They have also been used to replace the endogenous stem cells destroyed by high doses of chemotherapy and radiation used in cancer treatment (Saba et al. 2000). Because adult stem cells can be made to specialize into different kinds of cells, the generated cells could also be used to test the safety and efficacy of new drugs for a variety of diseases, thus removing the need for using animal models and shortening the drug development process. Scientists also envision using adult stem cells to grow complete organs such as the kidney and liver for autologous transplantation.

It must be emphasized that with the exception of the treatment for leukemia and other blood disorders, all other adult stem cell therapies are experimental in nature and do not therefore fit into the medical futility debate. It is not that adult stem cell interventions have a very low probability for success; rather, we simply do not have solid empirical evidence to support the medical use of adult stem cells at this point in time.

Ethical Issues

Adult Stem Cells and the Principle of Subsidiarity

Because the isolation of adult stem cells does not involve the destruction of embryos, some authors have invoked the principle of subsidiarity and called for a moratorium on human embryonic stem cell research in favor of adult stem cell research. According to this principle, the instrumental use of embryos can be justified only if no suitable alternatives exist for achieving the same goals (see Pennings and Steirteghem 2004; Town and Jones 2004). Indeed, opponents of embryonic stem cell research contend that such an alternative exists in the form of adult stem cells, which already have a track record of success in treating diseases. These cells have recently been found to be more versatile than previously thought, and because they can be taken from the patient whom they will treat, the problem of immune rejection can be circumvented. In the case of embryonic stem cells, this problem can only be avoided by using expensive immunosuppressive drugs or by cloning, which raises its own unique ethical problems.

A major problem with this line of reasoning, however, is that it assumes that the two research programs are alternatives. At the moment, scientists are unable to tell which of the two types of cells offers more potential for developing new therapies. Indeed, the general consensus within the scientific community seems to be that all stem cell research programs, including induced pluripotent stem cell (iPSC) research, should be pursued simultaneously in order to maximize the chances of discovering new therapies. Furthermore, the versatility of adult stem cells is still under debate, and compared to embryonic stem cells, they are less restricted in potency. This may limit how they can be used to treat diseases. Another point of concern is that adult stem cells are found in very small quantities and are difficult to access and culture.

Adult Stem Cell-Generated Gametes

Another significant ethical issue concerning adult stem cells is the possibility of creating human

gametes – spermatozoa and ova - from adult stem cells for reproductive purposes. A team of scientists has already been able to turn human bone marrow stem cells into spermatogonia, the precursor of sperm which can be grown into mature sperm (Drusenheimer et al. 2007). This technology could help gametically infertile couples and gay people (both male and female) to have their own genetically related children. This however raises a gamut of religious and moral quandaries which are unaddressed neither by the polemics of benefits nor the notion that both the donor and end user of adult stem cell-derived gametes are one and the same person. Besides, in order to test the functionality of gametes derived from stem cells, scientists will have to specifically create embryos for research purposes.

To be sure, those who believe that human embryos have the same moral status as adult humans might argue that this is tantamount to using another human being merely as a means to an end, which is morally objectionable. Moreover, gametes that are created in this manner have a high risk of being damaged or may have undetectable genetic abnormalities that could cause disease in offspring.

Stem Cell Tourism

But the main ethical issue concerning adult stem cells is the rise of stem cell tourism, a phenomenon where desperate patients from around the world travel abroad in search of unproven stem cell interventions not available in their home countries. Hundreds of stem cell clinics have sprouted throughout the world claiming to offer a cure for a host of neurodegenerative diseases ranging from muscular dystrophy, Stach's disease, and Alzheimer's disease. However, with the exception of stem cell transplants for certain types of cancers and blood disorders, there is no evidence to suggest that stem cell transplants are therapeutically efficacious. These clinics mainly operate in developing countries such as China, India, Thailand, Mexico, South Africa, the Dominican Republic, and Costa Rica. A few of these clinics are also to be found in America and Germany and in some eastern European countries such as Russia and Ukraine. The demand for stem cell interventions has been fuelled by a rapidly aging population and the concomitant increase in neurodegenerative disorders. The uncritical portrayal of stem cell therapies by the media has also contributed to the growth of this industry.

Most of these clinics are driven by commercial interests and have been accused of exploiting desperate patients (mainly from developed nations) by raising false hope and quick fixes. These clinics use only adult stem cells, which are marketed as a viable alternative to the ethically controversial embryonic stem cells. Adult stem cell therapies are also promoted as being natural and safe because they come from the patient's own body.

Unfortunately, the treatments being advertised by these clinics are mainly experimental in nature and lack clinical evidence of safety and efficacy. This puts patients at the risk of infection on the transplantation site, aberrant stem cell migration, tumor formation, neurological complications, and autoimmune rejection in case of allogenic transplantation. In addition, the targeted disease may still be present in case of autologous use. These unregulated treatments may also render the patients ineligible for genuine experimental treatments (Gunter et al. 2010). Additional medical risks to patients include inadequate postoperative care and monitoring as patients have to go back to their home countries after treatment. There is also concern that some of these clinics may not even be offering any stem cell interventions. Patients and their families are also at risk of experiencing financial harm. They have to raise funds to meet the high cost of treatment, travel, and accommodation, and some of these costs may not be covered by health insurance. The growth of these clinics also poses a threat to the field of stem cell medicine itself. Public confidence in cell-based therapies could be eroded, and scientists participating in legitimate clinical trials may be lumped together with the unscrupulous ones and denied research funds.

Access to Stem Cell Therapies and Global Justice

Another issue that needs to be addressed at this juncture is the positive and negative effects of

stem cell tourism on the health-care systems of the destination countries. Stem cell tourism is a lucrative source of revenue, which can be used to improve the public health infrastructure of the destination country. Indeed, stem cell tourism has been a huge source of foreign exchange for Asian economies.

But stem cell tourism (like all other forms of medical tourism) also raises questions of global justice. Although most stem cell clinics in developing countries such as India and China are run by locals, the majority of patients attending these clinics come from highly developed countries such as the USA and the UK. This creates internal brain drain in that although the local health-care workers do not physically leave their home countries, they attend to foreign patients. The local people are often too poor to afford the high fees charged by these clinics. It has also been argued that medical tourism (and this includes stem cell tourism) makes destination countries redirect resources away from basic health-care services, thus depriving and disadvantaging local citizens and distorting local health-care priorities.

Unless stem cell tourism is properly regulated, it has the potential of undermining efforts to address structural inequity in health-care provision in the destination countries. This can ultimately contribute to the perpetuation of global health inequalities.

Ethical Guidelines and Regulation

There have been attempts at both international and national levels to address the problem of unregulated stem cell therapies. Some of the measures that have already been taken include the drafting of international research guidelines for stem cell scientists, patient education, tightening of national regulation, and reinforcing the already existing ethical guidelines (Zarzeczny et al. 2014). The International Society for Stem Cell Research (ISSCR) has already developed and published *Guidelines for the Clinical Transplantation of Stem Cells* (2008a) and a *Patient Handbook for Stem Cell Therapies* (2008b) for those seeking stem cell-based interventions. In May 2011, the largest stem cell clinic in Europe called the Xcell-Center was closed down by the German authorities following the death of a baby undergoing autologous adult stem cell transplantation. The clinic had been operating through a loophole in the country's regulations concerning experimental treatments which has since been sealed. In the Netherlands, private stem cell clinics are prohibited (Sheldon 2007). Whereas some countries such as Turkey have no regulations for stem cell practices, a number of countries such as China, India, and Malaysia have developed their own stem cell guidelines and policies. In the USA, the Food and Drug Administration (FDA) has the authority to regulate adult stem cell therapies.

Of course regulatory guidelines are effective only to the extent they are recognized and obeyed. In some jurisdictions, stem cell research guidelines have proved difficult to enforce. This has largely been blamed on laxity and complicity on the part of the regulatory authorities especially government and medical establishments (Kiatpongsan and Sipp 2009).

Conclusion

Although adult stem cell research is not as controversial as embryonic stem cell research, it does raise ethical and policy challenges that are common to all types of stem cell research. The best way to address these challenges is to develop and implement a robust ethical and regulatory regime that will guarantee the safety and well-being of research participants. However, such regulation must be balanced against the need to promote scientific progress.

Cross-References

- ▶ Exploitation
- ▶ Moral Status
- ▶ Research: Human Subjects
- ▶ Safety, Patient
- ▶ Stem Cells: Embryonic
- ▶ Transplantation Medicine

References

Drusenheimer, N., Wulf, G., Nolte, J., et al. (2007). Putative human male germ cells from bone marrow stem cells. *Society of Reproduction and Fertility supplement, 63*, 69–76.

Grove, J. E., Bruscia, E., & Krause, D. S. (2004). Plasticity of bone marrow-derived stem cells. *Stem Cells, 22*, 487–500.

Gunter, K. C., Caplan, A. L., Mason, C., et al. (2010). Cell therapy medical tourism: Time for action. *Cytotherapy, 12*, 965–968.

ISSCR – International Society for Stem Cell Research. (2008a). *Guidelines for the clinical transplantation of stem cells*. Available at: http://www.isscr.org/clinical_trans/pdfs/ISSCRGLClinicalTrans.pdf. Accessed 10 Feb 2015.

ISSCR- International Society for Stem Cell Research. (2008b). *Patients' handbook on stem cell therapies*. Available at: http://www.isscr.org/clinical_trans/pdfs/ISSCRPatientHandbook.pdf. Accessed 10 Feb 2015.

Kiatpongsan, S., & Sipp, D. (2009). Monitoring and regulating offshore stem cell clinics. *Science, 323*(5921), 1564–1565.

Pennings, G., & Van Steirteghem, A. (2004). The subsidiarity principle in the context of embryonic stem cell research. *Human Reproduction, 19*(5), 1060–1064.

Ramalho-Santos, M., & Willenbring, H. (2007). On the origin of the term "stem cells". *Cell, 1*, 35–39.

Saba, N., Abraham, R., & Keating, A. (2000). Overview of autologous stem cell transplantation. *Critical Review of Oncology/Hematology, 36*(1), 27–48.

Sheldon, T. (2007). The Netherlands bans private stem cell therapy. *British Medical Journal, 334*(7583), 12.

Thompson, J. A., Itskovitz-Eldor, J., Shapiro, S. S. et al. (2008). Embryonic stem cell lines derived from human blastocysts. *Science, 2832*, 1147–1147.

Till, J. E., & McCulloch, E. A. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation Research, 14*, 213–221.

Town, C. R., & Jones, D. G. (2004). Stem cells, embryos and the environment: A context for both science and ethics. *Journal of Medical Ethics, 30*, 410–413.

Zarzeczny, A., Caulfield, T., Ogbogu, U., et al. (2014). Professional regulation: A potentially valuable tool in responding to "stem cell tourism". *Stem Cell Reports, 3*, 379–384.

Further Readings

Holland, S., Lebacqz, K., & Zoloth, L. (2001). *The human embryonic stem cell: Science, ethics and public policy.* Cambridge, MA/London: The MIT Press.

Munsie, M., & Hyun, I. (2014). A question of ethics: Selling autologous stem cell therapies flaunts professional standards. *Stem Cell Research, 13*(3), 647–653.

Stem Cells: Embryonic

D. Gareth Jones
University of Otago, Dunedin, New Zealand

Abstract

The advent of stem cells into the public arena in 1998 raised hopes for the treatment of a host of debilitating diseases affecting many organs and tissues of the body. The stem cells on which most attention was paid were embryonic stem cells, on account of their enormous biological potential. However, their derivation from human embryos raised major ethical qualms, involving as it does the destruction of embryos. The hype accompanying stem cells has had repercussions ranging from medical tourism to scientific fraud. The emergence in 2006 of induced pluripotent stem cells added to the clamor surrounding stem cells, since these appeared to surmount the ethical problems through bypassing an origin in embryos. These stem cells are obtained by reprogramming adult body cells so that they revert to a pluripotent state and appear to have a potential akin to that of embryonic stem cells. Embryos from which embryonic stem cells are derived may be nonviable, surplus embryos from IVF programs, embryos produced for research purposes, embryos created using somatic cell nuclear transfer, and human admixed embryos. Policy and regulatory frameworks governing the production of embryonic stem cells fall into four dominant categories, each with differing underlying ethical strictures and also with substantial scientific and clinical repercussions. The comparative status of embryonic and induced pluripotent stem cells is assessed both scientifically and ethically.

Keywords

Stem cells; Embryonic stem cells; Induced pluripotent stem cells; Embryo research; Moral status of the embryo

Introduction

It was in 1998 that human embryonic stem cells (ESCs) gained prominence when they were first successfully derived from human blastocysts, that is, early embryos (Thomson et al. 1998). Besides reproducing themselves sustainably and indefinitely in culture (self-renewal), these cells also maintain the developmental potential to form derivatives of all three embryonic germ cell layers. Such pluripotency enables them, under appropriate conditions, to become all the tissues in the body. These include gut epithelium (from endoderm); cartilage, bone, and muscle (from mesoderm); and neural epithelium and stratified squamous epithelium (from ectoderm). It is this latter property that is the distinguishing mark of embryonic (as opposed to adult) stem cells (Towns and Jones 2004). The potential of this discovery lay in its ability to produce large, purified populations of cells and neurons, particularly in diseases resulting from the death or dysfunction of one or a few cell types, such as Parkinson's disease and juvenile-onset diabetes mellitus. These prospects were accompanied by two sets of problems: ethical and the unrealistic expectations surrounding them.

The ethical problems arise because deriving stem cells from embryos entails their destruction, an act that immediately raises the contentious question of the moral status of embryos. For many, the use of ESCs is repugnant, as opposed to the use of adult stem cells (ASCs) that are regarded as ethically neutral. This, in turn, raises the question of the relative therapeutic efficacy of the two types. This is a scientific and clinical issue, although some who regard the use of ASCs as ethically preferable tend to conflate ethical and scientific arguments. This leads to highly contested claims that ASCs are therapeutically superior to ESCs. The result is a confusing mix of scientific, ethical, and theological considerations.

The debate on ESCs has been made far more problematic by the unrealistic expectations of the public, based on hype and exaggerations of the therapeutic potential of stem cells (ASCs as well as ESCs). This has led to a massive growth in stem cell tourism, as patients and their families search for the wonder cures promised by various clinics from the injection of stem cells. Unfortunately, most of these are not backed up by stringent peer-reviewed publications. This mixture of false hope and hype has tarnished the reputation of stem cell research, made worse by the excessive hype of legitimate researchers who sometimes make unrealistic claims about the short-term therapeutic benefits of stem cell advances in the clinic.

Public Face of ESCs

These and other intense pressures on stem cell researchers have had an even more tragic outcome, namely, the blossoming of pernicious scientific fraud in the stem cell arena. The major contributor to this dubious hall of fame was Hwang Woo-Suk who in 2004 appeared to have become the first scientist to clone human embryos and extract stem cells from them. This was followed in 2005 with the claim that his team had created the world's first ESCs using genetic material from patients and therefore matched to these patients. Later that year, he went further with the birth of the world's first cloned dog. The hype surrounding these breakthroughs was intense, but everything began to unravel when a series of ethical and scientific irregularities emerged, leading to serious questioning of the validity of the isolation of human ESCs (Cyranoski 2004, 2014). It was only far more recently, in 2013, that the production of patient-specific ESCs through cloning has been unequivocally accomplished (Tachibana et al. 2013).

Misconduct has also marred ASC research. Cardiologist Bodo-Eckehard Strauer claimed to have saved the life of a patient suffering from cardiogenic shock by transplanting adult autologous bone marrow-derived stem cells into a damaged artery. Described as a global innovation, the results based on a small number of cases have been severely critiqued on a range of fundamental errors, discrepancies, and contradictions (Francis et al. 2013).

Another striking debut for ESCs has been into the world of politics. In the United States,

President George W. Bush spoke to the nation on 9 August 2001 about ESC research, when he declared that "embryonic stem cell research is at the leading edge of a series of moral hazards." (Bush 2006) At that time, he announced that the use of NIH (federal) funds would be permitted for research on an estimated 60 stem cell lines already in existence as of that date. These lines must have been derived from embryos surplus to the requirements of IVF programs. No new embryos could be destroyed in deriving ESCs using federal funds. The aim of this dictate was to encourage respect for human life at the same time as exploring the promise and potential of stem cell research in finding cures for debilitating diseases. It is unfortunate that the stem cell lines already in existence, plus additional ones potentially eligible for federal research funding, failed to live up to ethical standards set by the Food and Drug Administration (FDA) (Jonlin 2014).

Emergence of iPSCs

A major breakthrough came in 2006 with the first description of induced pluripotent stem cells (iPSCs). This landmark study demonstrated that skin cells can be reprogrammed into stem cells. This was the first direct reprogramming of differentiated mammalian somatic cells back to a pluripotent state by transfecting the cells with four transcription factors (Takahashi and Yamanaka 2006). The resulting iPSCs appear very similar to ESCs and are also patient specific. Possible uses for iPSCs in human therapy include in vitro disease modeling (so-called disease in a dish), high-throughput drug discovery and screening, regenerative therapies, and even novel reproductive techniques.

The immediate response to these developments was positive, since they gave the impression of opening doors that had been shut on account of the ethical quandaries associated with ESCs and the destruction of embryos. Unfortunately, dubious and even fraudulent scientific studies were not far behind this Nobel Prize-winning work. A very short-lived episode hit the headlines in 2012 when Hisashi Moriguchi claimed to have cured six heart failure patients with cells derived from iPSCs. It soon emerged that these claims were baseless. In 2014, a major simplification of the iPSC technique developed by Haruko Obokata created international interest with the description of STAP (stimulus-triggered acquisition of pluripotency) cells. Together with coworkers, she had described how cells of various types, including skin, muscle, and lung cells, could be rapidly changed into an embryonic-like state by being dipped in a mild acid solution. However, issues quickly emerged over irregularities in images, suggesting at the least innocent mistakes and at the worst fraud.

Embryos and ESCs

ESCs are derived from the inner cell mass (ICM) of early embryos at the blastocyst stage, 5–7 days after fertilization. These ESCs have the ability to create all the cell lines of the embryo/fetus but not the individual itself. At present, their extraction disrupts the ICM and therefore destroys the blastocyst. The embryos used in this way have a number of sources (Jones and Whitaker 2009).

The first of these is nonviable embryos created via IVF. These will not be transferred to a woman since they are biologically incapable of further development. Use of these embryos is not contentious.

Second, and far more important in practice, is surplus embryos created during IVF programs. While created for implantation into a woman, they are no longer required for reproductive purposes. These embryos are viable, but unless donated to others in an IVF program will eventually be destroyed (allowed to thaw) since most legislation prohibits the indefinite frozen storage of surplus embryos. Since these embryos were created for reproductive purposes, it is possible to procedurally separate the decision to destroy surplus embryos from the decision to use them for research. This decreases the likelihood of exploitation and coercion.

The third source is embryos created specifically for research purposes. In this case, the destruction of embryos is premeditated, with

research as the only end point. There is no intention that the embryos will be allowed to develop into human beings. For many, this is a threat to human dignity, since it represents a further step in the instrumentalization and commodification of human life. However, societies that approve of procedures, such as IVF, prenatal diagnosis, preimplantation genetic diagnosis (PGD), and the creation, storage, and destruction of surplus embryos, give only limited respect to early embryos. Any differences between these procedures and those producing embryos explicitly for research purposes are ones of intention. However, in the research paradigm, embryos are being produced as a means to an end, and this sets this source apart from any others.

A fourth source takes the research goal further with the creation of embryos using somatic cell nuclear transfer (SCNT). The difference between this source and the creation of research embryos using IVF lies in the way in which the embryos are created. An argument against allowing SCNT (research cloning) is that it is the beginning of a "slippery slope" toward reproductive cloning and a devaluation of human life in general. A different concern is that SCNT could result in an improper use of women's bodies by creating a market for human eggs. This in turn may lead to the exploitation of poorer women, who would be the most likely to sell their eggs.

In an effort to combat the shortage of human eggs, a fifth source is that provided by human admixed (interspecies) embryos created for research purposes. This uses animal eggs to create a "cytoplasmic hybrid." The differences between embryos created via IVF or SCNT on one hand and the admixing of species on the other appear to be minor, although another biological boundary has been breached.

While this discussion has centered on the use of embryos for research, it is pivotal for the ESC debate, since ESCs can only be obtained from these embryos. Additionally, the embryos are in vitro blastocysts (those in the laboratory) and not in utero blastocysts (those in a woman's uterus and in an environment congenial to further development). While the latter have the potential of producing human individuals (totipotent), in vitro blastocysts have no such potential in the laboratory. Those on which research is conducted never acquire this potential, since research on human embryos beyond 14 days is currently forbidden.

Policy and Regulatory Frameworks Governing ESCs

Regulations governing ESCs fall into four dominant positions. These were designated A to D by Towns and Jones (2006). Position A encompasses countries that prohibit all embryo research and therefore the extraction of ESCs. Position B confines the use of embryonic stem cells to those currently in existence, in that they were extracted prior to a specified date, thereby prohibiting the extraction of ESCs and utilization of ESCs derived in the future. Position C allows for the use and ongoing isolation of ESCs from surplus IVF embryos from IVF programs. Position D allows the creation of human embryos specifically for research via both fertilization and SCNT. The Hinxton Group (An International Consortium on Stem Cells, Ethics and Law 2006) again identified four groups: Prohibitive (equivalent to A), Restrictive Compromise (B), Permissive Compromise (C), and Permissive (D). The classification adopted by the European Science Foundation (2013) is similar but omits a position B equivalent. The groups are Very Restrictive (corresponding to A), Permissive (C), and Very Permissive (D), with further categories of Restrictions by Default (where legislation is not explicit but national practices are quite restrictive in practice) and Unlegislated (where there is no legislation on human ESCs).

While some countries have moved between categories over recent years, the current situation is exemplified by the following examples:

A. (Prohibition): Italy, Slovakia, Tunisia
B. (Restrictive Compromise): United States – use of federal funds under President Bush
C. (Permissive [Compromise]): Iran, Saudi Arabia, China (Hong Kong), Taiwan, Canada, Denmark, France, Cyprus, Greece, Hungary, Iceland, the Netherlands, Norway, Portugal, Spain,

Switzerland, Australia, United States – use of federal funds under President Obama

D. ([Very] Permissive): United Kingdom, Singapore, Japan, Israel, Belgium, Sweden, South Korea, certain states in the United States using private funds

Restrictive by Default: Romania, Turkey, New Zealand

Unlegislated: Austria, Ireland, Luxembourg, Poland

Position A (Prohibition) exemplifies the stance that human life commences at fertilization, allowing nothing to be done to the embryo that is not in its best interests. Such a stance would also be expected to disapprove of IVF, the production of surplus embryos, and the derivation of ESCs from these embryos. Its emphasis is entirely on harm done to embryos, rather than on benefits that might accrue from research using ESCs. It neglects any interests beyond those of the very early embryo, including those with fertility problems.

The intention of position B (Restrictive Compromise) was to allow some research on human embryos, while aiming to protect embryos. This was achieved by allowing research only on stem cell lines already in existence, since the embryos from which these lines had been extracted had previously been destroyed. The destruction of any further embryos was forbidden. This compromise position took note of the plight of people with severe degenerating conditions who could, possibly, benefit from scientific advances (Towns and Jones 2006). However, these restrictive ESC guidelines fail to protect the large numbers of embryos destroyed daily by IVF procedures in fertility clinics.

Position C (Permissive [Compromise]) limits ESC research to surplus embryos from IVF programs. This allows both the utilization and extraction of new ESCs and eliminates arbitrary time limits on extraction. It accepts the destruction of already existing embryos no longer required within IVF programs. These in vitro blastocysts have no future as human individuals, since the decision has already been taken that they will not be donated to other individuals within an IVF program. This position therefore seeks to improve the health status of individuals suffering from common debilitating conditions, alongside providing early embryos with the care and respect due to human tissue.

Position D ([Very] Permissive) represents a dramatic moral shift since embryos are being created solely for research purposes, their creation being for the benefit of scientific research into developmental phenomena. As research subjects, the embryos do not benefit, neither are their interests taken into account. While the scientific exploration will probably have a therapeutic rationale, any claims made for this work are to be realistic. Justification is also needed why this research cannot be conducted on surplus embryos.

Human Admixed Embryos

Hybrid embryos (true hybrids) are those created by the fusion of gametes from human and nonhuman animals to produce an embryo which is a genetic mix of the contributing species. Cytoplasmic hybrids (cybrids) are created by performing SCNT to introduce a somatic cell from one species (e.g., human) into an enucleated egg from another. Cybrids allow the creation of stem cells from adult somatic cells without the use of human eggs. This enables stem cell lines to be derived from individuals with diseases that may subsequently be studied in the resulting stem cells. Chimeric embryos are created by inserting stem cells from one species into an existing embryo of another (e.g., mouse cells to human embryos). The aim is to produce particular types of stem cell lines or to examine how stem cells develop in the embryo.

The UK debate on the Human Fertilization and Embryology Bill brought the opposing arguments into the open (HFEA 2007). Scientists in favor of allowing the production of human admixed embryos argued that it will assist in the study of normal embryonic development and genetic disease, including a range of conditions such as motor neuron disease, Alzheimer's disease, Parkinson's disease, and some cancers. For supporters of the bill, such work is an inherently moral endeavor, since its aim is to harness the

potential of stem cell research for the benefit of human health. The arguments of opponents vary but include the unnaturalness of the procedures and the crossing of species boundaries. For some, they are morally repugnant and violate human dignity. While they may promote a mechanical view of the world, they are not devoid of moral boundaries. Moral repugnance is an unpredictable basis for moral judgments, although sentiments of disgust are deeply ingrained warning signs that alert us to moral wrongs. Species integrity may allow us to preserve a coherent, familiar moral terrain, although this has to be balanced against the prospective benefits held out by research of this nature.

Will iPSCs Replace ESCs?

The advent of iPSCs in 2006 was seen by many as a major breakthrough, not only on the scientific front but also for the ethical debate over the destruction of embryos to obtain ESCs. Initial responses by some commentators deemed them ethically unproblematic, the underlying premise being that iPSCs are scientifically very similar, or even identical, to ESCs. In spite of such assurances, especially by those opposed to the use of ESCs, there remain a series of practical and ethical considerations. Since this is a rapidly changing field, views will probably continue to undergo adjustment for some time to come, although some pointers are available (Bridge 2013).

Many scientific questions remain about both human iPSCs and ESCs, with considerable scientific disagreement regarding the safety and efficacy of the two cell types in future cell therapies. Consequently, most stem cell scientists do not consider that ESCs can be completely replaced by iPSCs. According to this view, ESCs remain the gold standard of pluripotency, and the goal of iPSC research is to achieve an ESC-like state. There is growing evidence that ESCs and iPSCs are not the same at an epigenetic and genetic level. Not only this, ESCs are currently considered to have greater therapeutic potential and to be much closer to being translated into a clinical setting than iPSCs. Consequently, further methodological and functional studies are needed to improve the reprogramming technique to generate iPSCs with therapeutic potential more akin to ESCs. This means that ESCs are still needed to understand the basic mechanism of pluripotency and self-renewal (see Bridge 2013 for details).

A wide array of ESC lines is needed for three reasons. First, the current ESC lines have significantly restricted ethnic diversity. Second, it is important that ESC lines are able to differentiate into the tissues of interest. If they are to be used in regenerative medicine, it is necessary for them to follow a desired lineage of differentiation. Third, in order to study human disease, ESCs need to be created that are disease specific. To date, ESCs representing a relatively small number of heritable diseases have been created, a repertoire that needs to be increased significantly. Successful modeling and development of treatments for genetic disorders would require the derivation of more ESC lines representative of specific human genetic diseases. Hence, instead of circumventing the moral problems associated with ESC research, it is becoming increasingly clear that far more work is required using ESCs as an important comparator for future iPSC research.

An unexpected future ethical dilemma associated with iPSC research would emerge if it becomes feasible to utilize human iPSCs to produce sperm and eggs. Were this to eventuate, these could theoretically be used to create embryos. While this lies in the future, the potential to produce iPSC-derived embryos would raise familiar questions about the moral status of these embryos. Even if this possibility never sees the light of day, it demonstrates that the emergence of iPSCs has not put an end to ethical deliberation.

As one looks to the future, there are at least three possible scenarios regarding the respective roles of ESCs and iPSCs (Solbakk 2008). The first is that iPSCs might completely replace ESCs. The second is that iPSCs will significantly reduce the number of ESCs needed by replacing them in certain types of research. The final scenario is that ESCs will remain central to the field of stem cell research. There has been a move away from ESC research, but this may reflect political and financial pressures, as much as scientific ones.

Patient safety, effectiveness for use in treatments, accessibility to large numbers of patients, and the moral status of ESCs and iPSCs will continue to dominate scientific and ethical discussion.

Conclusion

The prospects opened up by stem cells are enormous both scientifically and medically. Nevertheless, the ethical challenges are set to remain for the foreseeable future, since these will only disappear if the use of ESCs is replaced by ASCs or iPSCs. This looks unlikely, unless another major breakthrough occurs. Currently, there is no evidence that complete replacement of ESCs (and embryo destruction) will prove scientifically acceptable. This is not the last word ethically, although it strongly suggests that the debate over ESCs will continue unabated.

References

Bridge, S. (2013). Induced pluripotent stem cells: An alternative to embryonic stem cells? *Asian Bioethics Review, 5*, 25–39.

Bush, G. W. (2006). On stem cell research. In M. Ruse & C. A. Pynes (Eds.), *The stem cell controversy: Debating the issues* (2nd ed.). Amherst: Prometheus Books.

Cyranoski, D. (2004). Korea's stem-cell stars dogged by suspicion of ethical breach. *Nature, 429*, 3.

Cyranoski, D. (2014). Cloning comeback. *Nature, 505*, 468–471.

European Science Foundation. (2013). *Human stem cell research and regenerative medicine: Focus on European policy and scientific contributions*. Strasbourg: European Science Foundation.

Francis, D. P., et al. (2013). Autologous bone marrow-derived stem cell therapy in heart disease: Discrepancies and contradictions. *International Journal of Cardiology, 168*, 3381–3401.

Hinxton Group: An International Consortium on Stem Cells, Ethics and Law. (2006). World stem cell policies http://www.hinxtongroup.org/wp_ao_map.html. Accessed 28 May 2014.

Human Fertilisation Embryology Authority (HFEA). (2007). *Hybrids and Chimeras: A report on the findings of the consultation*. London: HFEA.

Jones, D. G., & Whitaker, M. I. (2009). *Speaking for the dead: The human body in biology and medicine* (2nd ed.). Farnham: Ashgate Publishing.

Jonlin, E. C. (2014). Differing standards fir the NIH stem cell registry and FDA approval render most federally funded hESC lines unsuitable for clinical use. *Cell Stem Cell*, *14*, 139–140.

Solbakk, J. H. (2008). The ethics of stem cell research: Can the disagreements be resolved? *Journal of Medical Ethics*, *34*, 831–832.

Tachibana, M., et al. (2013). Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell, 153*, 1228–1238.

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell, 126*, 663–676.

Thomson, J. A., et al. (1998). Embryonic stem cell lines derived from human blastocysts. *Science, 282*, 1145–1147.

Towns, C., & Jones, D. G. (2004). Stem cells, embryos and the environment: A context for both science and ethics. *Journal of Medical Ethics, 30*, 410–413.

Towns, C., & Jones, D. G. (2006). Stem cells: Policy and ethics. In M. Ruse & C. A. Pynes (Eds.), *The stem cell controversy: Debating the issues* (2nd ed., pp. 329–341). Amherst: New York, Prometheus Books.

Further Readings

Cohen, C. B. (2007). *Renewing the stuff of life: Stem cells, ethics and public policy*. New York: Oxford University Press.

President's Council on Bioethics. (2007). Advancing stem cell science without destroying human life, Domestic Policy Council, The White House.

Ruse, M., & Pynes, C. A. (Eds.). (2006). *The stem cell controversy: Debating the issues* (2nd ed.). Amherst: Prometheus Books.

Stewardship

Robin Attfield
Cardiff School of English, Communication and Philosophy, Cardiff University, Cardiff, Wales, UK

Abstract

The ancient origins of stewardship beliefs are traced to Plato, Neoplatonism, the Old Testament, and Christianity. Evidence of attitudes regarding human beings as entrusted to complete God's work of creation and nature as a trust is found in the Church fathers, as well as in the Reformation period. Traditionally the earth was regarded as belonging to the Creator, with humans as

custodians answerable for its care. Stewardship has often been understood in recent decades as a secular concept, with answerability owed to society. It need not be construed as having an anthropocentric basis. Indeed some definitions represent it as a secular and non-anthropocentric concept. Numerous ethical and political objections have been raised against it, and these are considered in turn and found unpersuasive. Stewardship involves a sense of justice and upholds the Precautionary Principle. It is applied here to biodiversity preservation, sustainable development, and climate change mitigation.

Keywords

Stewardship; Trustees; Responsibility; Answerability; Management; Justice; Precautionary Principle; Sustainable development; Climate change mitigation

Introduction

Stewardship involves being a trustee or guardian of goods such as time, money, or other resources and has in recent times been applied to the human responsibility for the care and management of the natural world (the focus of this entry). This theme can be traced back to Platonism and to the Bible. Stewards do not own what they are entrusted with, but are implicitly answerable to a higher authority or constituency, whether divine or human. Objections to stewardship are here addressed, and stewardship is related to the Precautionary Principle, to biodiversity preservation, to sustainable development, and to climate change mitigation.

History and Development

John Passmore traced belief in human responsibility for nature to Plato's *Phaedrus*, where it is said at 246b, "It is everywhere the responsibility of the animate to look after the inanimate" and subsequently to the Neoplatonist Iamblichus, who derived from this passage the view that humanity is sent to earth by God "to administer earthly things" and care for them in God's name (Passmore 1974, p. 28). While finding the language of stewardship in the New Testament, Passmore does not find it applied to the care of nature until the Reformation. But his key source, Clarence J. Glacken, takes the view that the Bible as a whole is supportive of human stewardship of nature (Glacken 1967, p. 168). *Pace* Lynn White, it is certainly not anthropocentric, as readers of Psalm 104 or of Job 38–41 will acknowledge, and in Romans 8 the whole creation is included in God's plan of salvation.

Passmore also discovers ancient origins for a related tradition, for which the role of humanity includes cooperation with nature with a view to its enhancement or perfection (rather than for human self-interest). This tradition he traces to the Stoic Posidonius (in the first century BCE) and to the Hermetic tract *Asclepius* (of the second century CE), which asserts that "God willed the Universe should not be complete until man has done his part." It then disappears, according to Passmore, until resuscitated by the German metaphysician Fichte. However, Glacken also finds what amounts to endorsement of cooperation with nature (albeit not under that name) in Church fathers such as Basil, Ambrose, and Theodoret and in the less well-known Cosmas Indicopleustes, as well as their medieval successors such as the Benedictines and Cistercians. A frequent image used by the Church fathers was the likeness of the created world to a house which has been left incomplete for humanity to furnish and adorn, completing the work of the Creator. This image is close enough to one of trusteeship or stewardship to show that stewardship beliefs were alive (and not in abeyance, as suggested by Passmore) during the period between Iamblichus and the Reformation (see further Santmire 1985).

Jean Calvin resuscitated the New Testament language of stewardship and related it to responsibility before God for the use of a person's time, talents, and possessions. Indeed Calvin's anthropocentrism has inclined some critics to accuse stewardship beliefs in general of being integrally anthropocentric. But this trait is less apparent in

the writings of Matthew Hale, a seventeenth-century Chief Justice who explicitly applied human stewardship to the world of nature. "The end of man's creation was, that he should be the viceroy of the great God of heaven and earth in this inferior world; his steward, *villicus* [farm-manager], bailiff or farmer of this goodly farm of the lower world." Only for this reason was man "invested with power, authority, right, dominion, trust and care, to correct and abridge the excesses and cruelties of the fiercer animals, to give protection and defence to the mansuete [tame] and useful, to preserve the species of divers vegetables (i.e., wild or domesticated plants), to improve them and others, to correct the redundance of unprofitable vegetables, to preserve the face of the earth in beauty, usefulness and fruitfulness." As Passmore comments, the farming of the earth by humanity is thus subject to preserving its beauty and refraining from degrading its resources: and for derelictions humanity can be called to account (Passmore 1974, p. 30). Hale grounds his remarks in the injunction of Genesis 2 to "dress and keep the garden," which many others have interpreted, like Hale, as a charter for stewardship of the world of nature.

Hale was not alone. Thus in 1713, Alexander Pope wrote that "The more entirely the inferior creation is submitted to our power, the more answerable we should seem for our mismanagement of it" (Attfield 1983, p. 43). This passage was written in criticism of vivisection, and Pope was followed by many of the British moralists in his opposition to cruelty toward animals. Meanwhile the attitudes of Church fathers such as Basil to nature were echoed by John Ray (in *The Wisdom of God Manifested in the Works of Creation* (1691)) in his depictions of how humanity can enhance the landscape, for example, with fruitful fields and orchards, and his themes soon became commonplaces for writers of natural theology. Thus, Christian writers readopted the theme of enhancing nature, well in advance of Romantic writers such as Fichte, in a manner that Hale would have recognized as advocacy of stewardship.

During the twentieth century, the themes both of Calvin and of Hale were widely taken up. Many Protestants used the metaphor of stewardship about accountability for the use of money, while many environmentalists applied it to responsibility for the care and preservation of nature. (While many environmentalists have retained a religious sense of stewardship, some have adopted a secular sense.) This entry concentrates on the latter of these uses and thus environmental stewardship, the use relevant to the conservation and preservation of the natural world.

Conceptual Clarification

Stewardship involves the responsible care or management of some good, together with some kind of answerability. While its sphere is sometimes regarded as time, talents, money, or even political power, for present purposes this is treated as the natural world or relevant parts of it. A recent variant (devised to avert charges of pretentiousness) treats it instead as concerned with the management of human behavior as it relates to the natural world (Welchman 2012). However, the various charges against stewardship will be considered in the sections on ethical aspects that follow.

Some commentators have claimed that the aims of stewardship are invariably human interests. But there is no restriction to human interests in the sense of stewardship conveyed in the previous paragraph, and historically there was no exclusive concern with human interests on the part of Plato, the Old Testament, the New Testament, the Church fathers, or writers such as Hale or Pope. Even ancient Stoics (such as Posidonius) seem not to have been exclusively anthropocentric, holding that everything exists for the sake of humanity *except the universe itself*. Some adherents of stewardship (of one kind or another) have been anthropocentrists, like Calvin, but many have not, and the possibility of adherents being biocentric and recognizing the moral standing of all living creatures remains an open one.

Others have maintained that in view of the implication of answerability, stewardship is essentially a religious notion, since if belief in answerability to God is discarded (or is merely

absent), then there is nobody for stewards to be answerable to. But this is a fallacy. Stewards cannot be answerable to future generations, as the latter will never be in a position to hold current agents to account, but they could be answerable to (say) the community of moral agents (past, present, and future), since the present segment of this community is capable of holding them to account, or simply to all their human contemporaries (for the same reason). Accordingly secular stewardship is a significant possibility, and this makes stewardship open to millions more people than the approach of the critics would suggest. Besides, if stewardship beliefs are to be found in Plato, then they will quite possibly have been secular beliefs, since Plato's theology fluctuated from belief in a demiurge (as in *Timaeus*) to an agnosticism associated with belief in the (uncreated) forms.

Stewardship beliefs are often (but not invariably) accompanied by denial that the earth is owned by the present generation of humanity. This denial can be expressed as in the opening of Psalm 24, "The Earth is the Lord's," but can be expressed in a secular manner, as in Karl Marx's rejection of the possibility that the present generation of humans possesses the earth for their exclusive benefit. It can also be expressed in the environmentalist commonplace that the earth does not belong to us, but that we borrow it from our children. While the latter claim cannot be literally true, the sense of responsibility that it seeks to convey makes room for some kind of recognition of stewardship, however anthropocentric.

A recent definition of stewardship is worthy of note: "Stewardship is the responsible use (including conservation) of natural resources in a way that takes full and balanced account of the interest of society, future generations, and other species as well as private needs, and accepts significant answerability to society" (Worrell and Appleby 2000, p. 269). This definition manifestly concerns secular stewardship of a non-anthropocentric kind and could for these reasons be construed as too narrow (through apparently excluding both religious versions and anthropocentric ones too), but these apparent defects serve to rectify the mistaken opposing claims that stewardship can never take either a secular or a non-anthropocentric form at all.

Traditionally, however, stewardship has been understood theocentrically, as in the following statement from the General Synod Board for Social Responsibility of the Church of England:

> Christians believe that this world belongs to God by creation, redemption and sustenance, and that he has entrusted it to humankind, made in his image and responsible to him; we are in the position of stewards, tenants, curators, trustees or guardians, whether or not we acknowledge this responsibility. Stewardship implies caring management, not selfish exploitation; it involves a concern both for present and future as well as self, and a recognition that the world we manage has an interest in its own survival and wellbeing independent of its value to us. Good stewardship requires justice, truthfulness, sensitivity, and compassion. It has implications . . . for individuals, organisations, and states. (Board for Social Responsibility 1991, p. 2).

Ethical Dimension

Stewardship has far-reaching ethical implications, both for biodiversity preservation, sustainable development, and climate change mitigation, but first several ethical, theological, and political objections need to be considered and then the Precautionary Principle, to which adherents of stewardship standardly subscribe.

Indelible Historical Associations?

The association of stewardship with long-standing theistic traditions, whether Jewish, Christian, or Islamic, has aroused objections that it is for these reasons an expression of a premodern hierarchical, oppressive, and/or sexist society. Those who deploy this kind of objection range from some Marxists, via some critical theorists, to some feminists, ecofeminists included.

To such objections, Jennifer Welchman well replies that we do not regard the comparable associations of democracy, which originated in sexist and slavery-dependent societies, as fatal to its modern acceptance, and that environmental stewardship has, like democracy, been significantly revised so as to outlive these historical

associations. Thus, contemporary stewardship has no links to oppression or to sexism and can be endorsed by both women and men alike. Her claims about modern revisions are borne out both by the definition of Worrell and Appleby and the statement of the General Synod Board of Social Responsibility (both quoted above): stewardship, as thus understood, does not regard human beings as ancient slaves or as medieval serfs nor women as subordinate, but treats men and women alike as free and responsible agents, entrusted with the care of the planet.

To ensure that undesirable historical associations do not attach to the stewardship of the twentieth and twenty-first centuries, adherents of stewardship are well advised to focus on some of the alternative metaphors included in the Board for Social Responsibility statement. Stewards are also trustees, entrusted with valuable goods, tenants expected to preserve the land, curators in charge of treasures to pass on to future generations, and guardians, whose charges have a value that is more than instrumental. To focus on these alternative metaphors can help rescue adherents of stewardship from any tendency toward entrapment in historical associations that they seek to disavow.

Devaluation of the Natural World?

This objection can take several forms, but they have in common the claim that the stewardship tradition separates God from the world of nature and thus prevents it from being respected. God, for the critics, is an absentee landlord. A theology such as pantheism, by contrast, confers a higher status on nature; if God were understood as coextensive with the world, then more salutary ethical practices would be promoted – or so it is sometimes suggested (see Palmer 1992). This is the kind of objection inspired by White's aspersions on Judeo-Christian theology (White 1967).

But the doctrine of creation, which certainly advocates worship of the Creator and not of creatures, at the same time involves regarding the world as an expression of God's creative purposes, and God as indwelling the world (rather than absent from it). It requires human beings to respect nature as God's creation and other creatures as fellow creatures. While for pantheism God is material and there is no Creator independent of nature itself, to whom worship and service are owed, theism is itself consistent with regarding God as present in his or her creation, and in the forms of Judaism, Christianity, and Islam advocates, for example, the preservation of natural species and their habitats as creatures of God.

While the Bible confers on humanity dominion over nature (Genesis 1; Psalm 8), dominion is misinterpreted if it is taken to authorize domineering or human domination. Indeed the commandment to take care of the garden in which Adam and Eve were placed (Genesis 2:15) can instead be understood as requiring nothing less than stewardship (see further Attfield 2000).

A Pre-evolutionary View?

Palmer (1992) has also suggested that the stewardship model represents humanity as God's deputy on earth, for whom, at least in some versions, everything was made, whereas, if we accept Darwinism, humanity is as much a product of evolution as other species are, and the survival of those species involves not human government but their being left alone.

However, though the view that everything was made for the sake of humanity has sometimes been held, it conflicts with many Biblical passages (see above). Thus, passages like Psalm 104 and Job 38–41 presuppose that nonhuman creatures have an independent place in the Creator's intentions. There is indeed no inconsistency between theistic belief and acceptance of Darwinism, including its implications about human origins and vulnerability, and human survival not being guaranteed; and theists can recognize that stewardship often requires nonintervention, as well as planning the use of natural resources so that both present and future needs (human and nonhuman) can be satisfied. Secular adherents of stewardship are, of course, free to agree (see further Attfield 2003).

So far removed are stewardship beliefs from embodying a pre-evolutionary view that some secular advocates of stewardship have actually suggested that it could involve humanity in taking charge of the direction of evolution through large-

scale genetic engineering (Attfield 1999). Such measures, however, fall foul of the Precautionary Principle (see below), which adherents of stewardship characteristically uphold.

Stewardship as Managerialism?

A range of critics have alleged that stewardship involves human interference with the entire surface of the planet in order to enhance the productivity of nature's resources. Stewardship thus stands charged with an instrumentalist attitude to nature. It has also been claimed to involve a managerial model (as in the role of an ancient household steward), for which interventionism is a natural corollary. Yet human capacities are inadequate for such a managerial role, and, as James Lovelock claims, the planet has no need for such human management (Lovelock 2006). Palmer's conclusion is that "Stewardship is inappropriate for some of the planet some of the time, some of it for all of the time (the deep oceans), and all of it for some of the time – that is, before humanity evolved and after its extinction" (Palmer 1992, p. 79).

But there is no need for adherents of stewardship to adopt an instrumentalist attitude to nature, particularly when many Biblical passages appear to recognize its intrinsic value. And recognition of this value involves respect for other species and their habitats and thus refraining from colonizing the entire surface of the planet (rather than an approach of cosmic management). As we have already seen, there is no need for all the historical associations of ancient stewardship to be endorsed by modern stewards, and in any case stewardship is far from synonymous with interventionism.

Indeed stewardship is compatible rather with letting-be, appropriate for Palmer's own example of Antarctica. And while Palmer is right in holding that there was no human responsibility before there were human beings, and that there will be none after human extinction, responsibility remains possible for the entire sphere of nature which humans can affect, which, in the twenty-first century, includes, for better or for worse, the deep oceans, the solar system, and much of the outer space beyond it. Unless the correspondingly extensive human power is exercised with responsibility, global problems will be intensified. Thus, far from stewardship being inappropriate for any of the spheres of human activity, modern technology actually makes an attitude like stewardship indispensable (Attfield 1999, p. 55).

Further, Bruce Reichenbach and V. Elving Anderson aptly reply to Lovelock that the very arrival of humanity on the planetary scene, and thus of human disruption, is what makes stewardship both possible and necessary (Reichenbach and Anderson 2006). This defense of theistic stewardship, indeed, is just as appropriate as a defense of secular stewardship. What stewards are doing is mitigating the human footprint and ameliorating its ecological impacts. The related indispensability of stewardship suggests that, far from its practice being arrogant, its non-practice could be seen as amounting to negligence.

Reductionism?

Lovelock further supposes that stewardship embodies a form of "reductionism" that disregards the self-regulation that may, in his view, already be exercised by the superorganism "Gaia." This charge may seem to be a methodological one but turns out in practice to be an ethical objection. Stewards, he believes, will be prone to reach for technological solutions such as geo-engineering to solve the problem of climate change, favoring, for example, saturating the oceans with iron chloride to fix surplus carbon dioxide through the growth of algae. But this "gunboat diplomacy" approach strongly conflicts with recognizing the Precautionary Principle (see below) which environmental stewards are both free and prone to favor, and suggests that Lovelock confuses stewardship and the technology of neocolonialism. Nor can we rely on planetary systems, however self-regulating they may be, to curtail either anthropogenic climate change or rapid species loss, for both of these dire processes are advancing despite the planetary systems that are in place.

Perhaps this is why Lovelock goes on to advocate seeing ourselves as planetary physicians, instead of stewards (Lovelock 2006, pp. 106–111), taking steps to protect vulnerable species and mitigate greenhouse gas emissions.

But if the misunderstandings about reductionism and the supposed arrogance of stewardship could be set aside, there would be no reason why we should not see ourselves in both these roles (planetary physicians and also stewards) at the same time, as long as the role of planetary physician is not construed as involving planet-wide interventionism and remains, like stewardship, consistent with widespread letting-be.

Neglect of Social Justice?

Yet other critics of stewardship have maintained that it is liable to ignore social and international justice and focuses instead on managing time, talent, and treasure, albeit sometimes in the name of the kingdom of God. Maybe some adherents of stewardship are tempted in this direction, but if so, they would be falling into a different kind of reductionism, in which the ethical basis of stewardship would largely be ignored, for the sake of a limited focus on the resources over which stewardship is most immediately exercised.

But that is not the stance of Worrell and Appleby, whose secular definition of stewardship requires stewards to take fully into account "the interest of society, future generations, and other species as well as private needs," and also "accept significant answerability to society." Nor is it the stance of the General Synod Board for Social Responsibility, whose statement, quoted above, involves stewards having "a concern both for present and future as well as self, and a recognition that the world we manage has an interest in its own survival and wellbeing independent of its value to us," adding that "Good stewardship requires justice, truthfulness, sensitivity, and compassion."

For both these definitions, then, stewardship involves a concern for justice, and thus for the poor and disadvantaged, and for developing countries as well as developed ones, for future generations, and for nonhuman species and their interests both in the present and in the future. This broad and deep ethical basis cannot be relinquished if stewardship is to remain true to itself. Besides, stewardship involves not just responsibility but also answerability, an acceptance which makes delivery of the concerns just mentioned far more likely to be taken seriously. As was mentioned above, answerability is sometimes seen as owed to God, but can also be understood as owed to the community of moral agents or, as the definition of Worrell and Appleby advocates, "to society."

As a broad ethical platform, the stance of stewardship is neutral between the various forms of normative ethics. Thus, it can be upheld by deontologists, by consequentialists, by virtue ethicists, and by rights theorists (particularly those who emphasize animal rights). It can also advocate the cultivation of virtues, grounding these not on a virtue ethics basis but on a basis of rights, consequentialism, or Kantianism (see further Attfield 2012).

Nevertheless there would be a problem of justice if everyone were to be treated as having the same degree of responsibility as everyone else for the care of the environment and the natural world. People living from hand to mouth, however, cannot be expected to make provision for future generations when their own day-to-day survival is itself precarious. Individuals and communities vary enormously in their powers and capacities, as also do corporations and countries, and both responsibility and answerability vary accordingly. Hence often the stewardship of the poor cannot be put into effect because of their own lack of resources, while much greater responsibility rests with those who wield financial and/or political power. Sometimes, certainly, schemes and policies are devised which encourage local people of limited means to preserve their own environments, forests, and wildlife, and such schemes would appear greatly preferable to policies of excluding forest peoples from their own forests. Yet the requirements of justice extend further. To the extent that those committed to stewardship are also committed to justice, they will look for ways of enhancing the agency of the poor, such that they too can participate in the stewardship of the environment, which committed adherents of stewardship recognize as the role not only of themselves but also of humanity.

An Establishment View?

The sheer number and variety of the objections to an apparently salutary ethical stance of itself calls

for diagnosis. An analogy is the concept of sustainable development (see below), an equally salutary concept, which in addition carries the unanimous support of all the countries represented at the Rio Summit of 1992 and at subsequent Summits on environment and development. In the case of sustainable development, objections can widely be traced to the strong desire on the part of these countries (subsequent to 1992), and on the part of the companies which seek their support, to reinterpret sustainable development in ways that align it with their own policies, and this has led to accusations to the effect that "sustainable development" has come to mean "business as usual." In fact, though, it remains a radical concept concerning the satisfaction of both current and future needs, which should not be discarded just because lip service is so widely paid to it. For if this were a reason to reject it, we should equally reject democracy as well.

In the case of environmental stewardship, there has been equally widespread support of this concept, all the easier to adopt in view of "stewardship" being fundamentally a metaphor with ethical overtones that are difficult to resist. The term "stewardship" has accordingly been used not only by religious bodies but also by governments and by financial institutions, to such a degree that skeptics find it easy to represent it as a cliché meaning little more than "business as usual" or as a trite expression of tired establishment ethics. But this view of its meaning involves a further form of reductionism, curable by reverting to definitions of stewardship such as that of Hale or that of Worrell and Appleby. The application of human ethical responsibility to our environmental problems remains crucial, together with acceptance of answerability, whether to God or to the moral community, and that is precisely what the concept environmental stewardship offers (and what the critics could join in recognizing).

The Precautionary Principle

The Precautionary Principle concerns the avoiding of harm that is either irreversible or serious and reversible but only with great difficulty and great effort. Because irreversibility is much more obviously a feature of environmental resources than cultural ones, it is to these that the Principle is most often applied. The Principle declares that where there are threats of serious or irreversible damage (environmental damage included), lack of full scientific certainty or knowledge should not be used as a reason for postponing measures to prevent this damage. It thus transcends principles which seek to prevent damage once a risk has been established, and concerns cases of uncertainty, where the probability of damage cannot be predicted, but where there is reason to believe it likely.

The attitude of environmental stewards to this Principle is almost invariably one of acceptance, for it safeguards those resources of which stewards see themselves as custodians, advocating action to avoid serious or irreversible damage. While some object that this Principle advocates excessive caution, this objection is itself based on the misunderstanding of conflating the Precautionary Principle with the Principle of Maximin, which advocates selecting the option (inaction included among options) of which the worst conceivable outcome would be least bad. But the Precautionary Principle is not concerned with outcomes that are merely conceivably possible in theory, but ones which there is reason to credit. Accordingly there is nothing to prevent its adoption by adherents of stewardship.

The same reasoning implies that policies of stewardship can be expected to comply with this Principle. This is why adherents of stewardship will usually steer clear of "technological fixes" (see above) such as the more radical forms of geo-engineering. They will not avoid all forms of modern technology, because in some circumstances applications of, for example, genetic engineering could be crucial in averting famine. But where technology embodies serious or irreversible risks, as with most forms of solar radiation management, they will advocate other policies instead, such as ones of mitigation and adaptation (see below).

Biodiversity Preservation

Faced with a loss of species of an almost unprecedented kind, adherents of environmental stewardship will support measures of preservation and in some cases restoration. Where, for example,

plantations in Indonesia have illegally been extended along the very banks of rivers, depriving wildlife of riparian corridors sufficient to sustain the viability of local populations, they will support the rewilding of riverside stretches, either through the regrowth of forest after plantations have been pulled back or through the deliberate planting of forest species. These, after all, are policies that the Precautionary Principle supports (further instances of its capacity to advocate intervention rather than cautious passivity).

The same people can consistently support the international agreement on biodiversity preservation made at Nagoya (Japan) in 2010. International collaboration is going to be vital if biodiversity-rich countries, which are often developing countries, are to receive the support they need to preserve biodiversity from less biodiverse countries which often have greater resources. While stewardship does not, as such, prescribe particular national policies, let alone international ones, it can supply grounds for positive action at all levels, including these.

Another example is here in place. In an address delivered at Cardiff University in 2013, Fazlun Khalid related that as soon as Qur'anic insights about responsibility for the environment were translated into Swahili in 2001 and conveyed to the fishermen of Zanzibar, they immediately abandoned their long-standing practice of dynamiting coral reefs; disobedience to the state was one thing, but disobeying Allah was quite another (see further Khalid and O'Brien 1992).

Sustainable Development

Sustainable development involves provision for present needs in ways that, far from undermining provision for future needs, put in place systems and policies that facilitate their fulfillment. As mentioned above, sustainable development was granted approval by nearly 200 countries at the Rio Summit of 1992 and at subsequent international conferences on environment and development.

Those who accept the definition of environmental stewardship of Worrell and Appleby will notice that, in matters of the treatment of natural resources, that definition commits them to policies of sustainable development in this sphere. Adherents of environmental stewardship need not be committed to sustainable development for all spheres (such as those of population), but will be guided by the requirements of the preservation of some resources and the conservation of others for use by future generations when it comes to policies for, for example, energy generation. Thus, there is a strong non-accidental link between stewardship on the one hand and policies of sustainable development on the other.

Climate Change Mitigation

Similar implications arise in connection with climate change. If resources such as clean air, a viable climate, and intact ecosystems (forests, wetlands, and coral reefs included) are to be available to successive future generations, then urgent steps are needed to mitigate anthropogenic climate change, and also to adapt to the emissions of carbon dioxide and other greenhouse gases that have already taken place. But just such preservation of resources is required by stewardship, if understood along the lines of the Worrell and Appleby definition or that of the Board for Social Responsibility.

When we also bear in mind that to avert an average increase in global temperatures of over two degrees (Celsius) humanity can emit no more than a trillion tonnes of carbon dioxide (or equivalent) across time, and that well over half of this total has been emitted already (Attfield 2014), the urgency of an international agreement on greenhouse gas mitigation is underlined. Forms of adaptation, such as seawalls and flood barriers, are also essential, but without mitigation the underlying problems will become increasingly intractable. Advocacy of such an agreement thus turns out to be a natural application of commitment to environmental stewardship.

Conclusion

The religious and theistic tradition of stewardship well equips its adherents to uphold an ethic of environmental concern, not least for the conservation of resources that future generations can be

foreseen to need, and for the preservation of as many as possible species, habitats, and ecosystems. However, adherence to such an ethic does not depend on allegiance to any of the religions, for secular versions of stewardship can be embraced, recognizing answerability to society or to the community of moral agents, rather than to God.

This ancient tradition has been applied across the centuries of the modern period to the treatment of nonhuman animals and to human interactions with the natural environment. It is arguably immune from charges of commitment to oppressive social practices of the ancient and/or medieval periods, of devaluing nature, of embodying a pre-evolutionary view or managerialism, of commitment to technological reductionism, or of neglect of social justice, or again of upholding an establishment stance of "business as usual." Its adherents are likely to be committed to a range of ethical values and principles such as the Precautionary Principle and to support biodiversity preservation, sustainable development, and climate change mitigation. These considerations suggest that it may well supply a viable and tenable option for addressing many of the problems and vicissitudes of the contemporary world.

Cross-References

- ▶ Agricultural Ethics
- ▶ Bioethics: Environmental
- ▶ Food Ethics
- ▶ Future Generations
- ▶ Justice: Global
- ▶ Moral Theories
- ▶ Precautionary Principle
- ▶ Responsibility: Individual
- ▶ Responsibility: Social
- ▶ Sustainability
- ▶ Trust

References

Attfield, R. (1991). *The ethics of environmental concern* (2nd ed.). Athens/London: University of Georgia Press.

Attfield, R. (1999). *The ethics of the global environment*. Edinburgh/West Lafayette: Edinburgh University Press/Purdue University Press.

Attfield, R. (2000). Christianity. In D. Jamieson (Ed.), *A companion to environmental philosophy* (pp. 96–110). Oxford: Blackwell.

Attfield, R. (2003). *Creation, evolution and meaning*. Aldershot/Burlington: Ashgate.

Attfield, R. (2012). *Ethics: An overview*. London/New York: Continuum/Bloomsbury.

Attfield, R. (2014). *Environmental ethics* (2nd ed.). Cambridge, UK/Malden: Polity.

General Synod Board for Social Responsibility. (1991). *Christians and the environment*. London: Board for Social Responsibility.

Glacken, C. J. (1967). *Traces on the Rhodian shore: Nature and culture in Western thought from ancient times to the end of the eighteenth century*. Berkeley/London: University of California Press.

Khalid, F., & O'Brien, J. (Eds.). (1992). *Islam and ecology*. London/New York: Cassell.

Lovelock, J. (2006). The fallible concept of stewardship of the Earth. In R. J. Berry (Ed.), *Environmental stewardship* (pp. 106–111). London/New York: T. and T. Clark.

Palmer, C. (1992). Stewardship: a case study in environmental ethics. In J. Ball, M. Goodhall, C. Palmer, & J. Reader (Eds.), *The Earth beneath* (pp. 67–86). London: SPCK; reprinted in R.J. Berry (ed.), *Environmental Stewardship*, London and New York: T. and T. Clark, 2006, 63–75.

Passmore, J. (1974). *Man's Responsibility for Nature*, London: Duckworth.

Reichenbach, B., & Anderson, V. E. (2006). Tensions in a stewardship paradigm. In R. J. Berry (Ed.), *Environmental stewardship* (pp. 112–125). London/New York: T. and T. Clark.

Santmire, H. P. (1985). *The travail of nature: The ambiguous ecological promise of Christian theology*. Philadelphia: Fortress Press.

Welchman, J. (2012). A defence of environmental stewardship. *Environmental Values, 21*(3), 297–316.

White, L., Jr. (1967). The historical roots of our ecological crisis. *Science, 155*(37), 1203–1207.

Worrell, R., & Appleby, M. C. (2000). Stewardship of natural resources: definition, ethical and practical aspects. *Journal of Agricultural and Environmental Ethics, 12*, 263–277.

Further Readings

Attfield, R. (2015). *The ethics of the global environment* (2nd ed.). Edinburgh: Edinburgh University Press.

Berry, R. J. (1995). Creation and the environment. *Science and Christian Belief, 7*(1), 21–43.

Warren, K. J. (1994). Environmental stewardship: An ecofeminist philosophical perspective. In M. A. Beavis (Ed.), *Environmental stewardship: History, theory and practice – Workshop proceedings, 11–12 Mar 1994*, (pp. 42–58). Winnipeg: University of Winnipeg.

Stigmatization

Justin Chen[1] and Andrew Courtwright[2]
[1]Depression Clinical and Research Program, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[2]Department of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital and Institute for Patient Care, Massachusetts General Hospital, Boston, MA, USA

Abstract

This entry reviews the definition, public health consequences, and moral status of stigmatization. Stigmatization involves identifying and marking an undesirable characteristic in a way that narrows a person's social identity to that characteristic. The consequences of stigmatization include marginalization and, in some cases, dehumanization. Stigmatization often contributes to poor global health outcomes, particularly for the diagnosis and treatment of infectious diseases and mental illness. In other cases, however, such as smoking cessation, stigmatization may result in improved health outcomes. Both consequentialist and non-consequentialist frameworks address the ethics of using stigmatization as a public health tool although these theories reach different conclusions.

Keywords

Denormalization; Public health ethics; Social identity; Stigma

Introduction

The term stigma is derived from the Greek stem *stig-* (mark or tattoo) plus *ma* (which denotes an action). Taken literally, a stigma is a mark. The sorts of persons and activities that have been historically stigmatized, however, make it clear that stigmatization is not a neutral act. For example, Greek amphorae from the fifth century BC represent injustice as a figure covered in marks associated with criminal acts. Branding, rather than tattooing, was a common method of marking criminals in the Byzantine Empire under Constantine I and well into eighteenth century France, where the courts could order prisoners branded with the *fleur-de-lis*.

Historically, stigmatization extended beyond criminal acts to other physically and morally undesirable traits, including contagions. Medieval lepers insufficiently marked by the pox of the disease had to wear a bell to warn others to keep their distance. Nineteenth-century prostitutes with cutaneous manifestations of syphilis were forced into infirmaries designed specifically to segregate syphilitics from the general hospital population. So-called moral contagions – such as pregnancy out of wedlock, as represented in Hawthorne's *The Scarlet Letter*, or mental illness, as depicted in Francisco Goya's paintings of asylums – were also common targets of stigmatization. Unwed mothers and the mentally ill were often institutionalized to separate them from the broader community, at least until their condition had resolved.

While not as significant a social force, anti-stigmatization campaigns have strong historical roots in the Christian and Enlightenment traditions. Drawing on New Testament concerns for prostitutes, tax collectors, Samaritans and other non-Jewish ethnic groups, and lepers, early Christian movements embraced radical equality with stigmatized populations. From a secular standpoint, William Shakespeare's portrait of Shylock ("Hath not a Jew eyes?"), Victor Hugo's sympathetic description of Quasimodo in *The Hunchback of Notre Dame*, and William Defoe's focus on illegitimacy and determinism in *Moll Flanders* can all be interpreted as historical literary efforts to destigmatize certain populations.

Stigmatization and Global Health

Despite anti-stigmatization efforts, however, stigma continues to play an active and largely destructive role in global public health. Though subject to cultural variation, stigmatizing beliefs are attached to a wide range of health conditions,

including schizophrenia, depression, sexually transmitted diseases, epilepsy, physical disability, leprosy, and drug, alcohol, and tobacco addiction. Epidemiologists and social scientists have found that stigmatization contributes to preventable morbidity and mortality by driving harmful behaviors underground, delaying care seeking, and decreasing life chances. Self-stigma, or the internalization of stigmatizing societal beliefs, has also been linked to lower self-esteem and non-adherence to recommended treatment, both of which exert a negative impact on the course of an illness (Corrigan et al. 2009). Increased stress associated with stigmatization also directly affects health through neurohormonal activation, leaving stigmatized populations at increased risk for cardiovascular and endocrine diseases (Marmot 2006).

Stigmatization also exacerbates health disparities both because it tends to affect groups already at risk for worse health and because marginalized groups are more likely to have stigmatizable conditions.

For example, among individuals with acquired immunodeficiency syndrome (AIDS), the experience of stigmatization in the form of negativity, discrimination, and social avoidance is associated with increased rates of depression, decreased clinic attendance, and greater reluctance to disclose human immunodeficiency virus (HIV) status to potential partners. Among HIV-positive patients, a high degree of perceived stigma is associated with poor adherence to antiretroviral therapy. Perceived stigma is highest among those HIV-positive patients – especially the poor and uneducated – who are already at risk for health disparities. Similarly, the stigmatization of tuberculosis (TB) is associated with diagnostic and treatment delay, particularly in populations already at risk for health disparities including immigrants, women, and the poor. TB-associated stigmatization has significant economic consequences, including job loss and exclusion from market spaces, and decreased social opportunities, particularly marriage prospects for single women. In the developing world, TB is also perceived as a marker for HIV status, resulting in the transfer of HIV-associated stigma to individuals with TB.

Stigmatization similarly impedes efforts to identify and treat mental illness. For example, immigrant minority women who perceive that their community stigmatizes depression are less likely to seek treatment for depressive symptoms. Among Chinese men and women with schizophrenia, self-stigma is the single strongest predictor of treatment compliance, accounting for 60 % of the variability in clinic attendance. Similarly, stigma is a strong predictor for treatment-seeking behavior for individuals with depression and schizophrenia across multiple socioeconomic groups.

Among obese individuals, higher rates of stigma are associated with increased body dissatisfaction, decreased self-esteem, and increased exercise avoidance, and obesity negatively impacts physician attitudes toward their patients. Other conditions in which stigmatization has been found to play a role in health and healthcare-seeking behaviors include: intravenous drug use, smoking-related cancers such as head and neck and lung cancer, fibromyalgia, urinary incontinence, and various dermatologic conditions including psoriasis and eczema.

As this brief review suggests, much of the research on stigma's health effects has examined single outcomes at one level of analysis – for example, associations between stigma and self-esteem among individual respondents with a particular illness (Hatzenbuehler et al. 2013). Researchers have carried out parallel track investigations for specific disease categories, including mental illness, obesity, HIV/AIDS, disability, and non-pathological but stigmatized characteristics including minority sexual orientation and race/ethnicity. They have examined a similarly diverse range of outcomes, including social and economic factors (e.g., housing, employment, education, and social relationships), psychological/behavioral consequences, and specific measures of health. As a result of this fragmentation, much of the current research does not treat stigma as an important unifying construct with implications for the social determinants of population health. As the psychologist Mark Hatzenbuehler (2013) and colleagues argue, however, stigma should itself be considered a fundamental cause of health inequalities.

Stigma may also manifest differently depending on cultural interpretations of illness. For example, stigma regarding mental illness in

China spreads quickly from the affected individual to his or her family because of etiological beliefs regarding mental illness that assign a "moral defect" to sufferers and their families (Yang et al. 2007). This contagion model results in a kind of "social death that threatens the very existence, value, and perpetuity of the family group" for individuals with mental illness and their relatives (Yang et al. 2007, p. 1529). In Japan, by contrast, mental illness is often considered to be a weakness of personality from which a person can never recover, and the majority of the general public maintains a greater social distance from affected individuals.

To better assess the role of stigma in health outcomes, further research is needed to understand how stigma is influenced by broader sociocultural factors such as education, immigration, urbanicity, income, and religion. It is also important to clarify whether stigma regarding different disease processes (e.g., schizophrenia vs. HIV/AIDS vs. physical disability) is directly comparable given wide variations in beliefs regarding the etiology and the social significance of each disease.

Defining Stigmatization

Despite the well-acknowledged impact of stigmatization on global health, social scientists, historians, epidemiologists, and bioethicists do not uniformly agree on a definition of stigmatization. This is partly because stigma scholars have focused on different aspects of stigmatization, including the motivations for social groups to stigmatize, the health and economic effects of stigmatization, why some people or traits are stigmatized and not others, and the justifications and social norms that promote or allow stigmatization.

The most commonly used model of stigmatization comes from the sociologists Bruce Link and Jo Phalen (2001). Ultimately grounded in social, economic, and power structures, they argue that stigmatization comprises a fourfold process: distinguishing differences between persons, linking those differences to negative stereotypes, creating social distance based on the marked trait, and losing status with consequent discrimination. For example, a person who has been institutionalized for mental illness is marked because *having had a mental illness* is associated with the negative stereotype *dangerous*. As a result, he or she becomes isolated and suffers status loss and discrimination and, subsequently, fewer social and economic opportunities.

Simplifying Link and Phalen's approach, there are two essential components of stigmatization: first, a trait, activity, or characteristic is identified and marked as undesirable; and second, the stigmatized person suffers a characteristic set of consequences, which include narrowing his or her social identity to the marked trait. Regarding the identification of a particular trait, this occurs when the stigmatizer judges, in accordance with community norms of desirability, that he or she would not want to have the trait himself or herself, that it should be removed from the community, and that the stigmatized person should also want to be free of it. As Link and Phalen suggest, this judgment often involves negative stereotypes but can also involve community standards of beauty or virtue or judgments about threats to community health or welfare.

The marking of a stigmatized trait can take the form of a physical disfiguration such as branding but more commonly involves behavioral or attitudinal changes toward the stigmatized person. Common attitudes in this context include contempt and disgust and common behaviors include avoidance or evasion, shunning, and even institutionalization or incarceration. The physical, behavioral, and attitudinal changes that mark the identified trait as undesirable result in a narrowing of social identity, a characteristic consequence of stigmatization. The sociologist Erving Goffman (1963) describes this as a "spoiled identity," and writes that stigmatization transforms an individual "from a whole and usual person to a tainted, discounted one." Expanding on this theme, the philosopher Martha Nussbaum argues that stigmatization reduces a person's social identity to only the marked trait, representing a "loss of uniqueness: the offender becomes a member of a degraded class" (Nussbaum 2006a). If the marked trait is the primary focus of an individual's social interactions, this prevents him or her from being

seen as a human being with a complex social identity and interests.

As the psychologist Patrick Corrigan notes, a spoiled or narrowed social identity is also perpetuated and reinforced through self-regarding attitudes. Looking beyond Link and Phalen's model, which focuses on stigma largely as an other-regarding process, the stigmatizer does not merely respond to a trait that is socially undesirable but demands that the stigmatized person share his or her judgment that the trait is undesirable. For example, when a person with obesity is stigmatized, the expected consequence is not just social isolation but that obese person also feels ashamed for being obese. In this way, stigmatization is closely associated with shame and self-loathing such that the stigmatized person is encouraged to hide the marked aspect of him or herself because it is disgusting. Stigmatization can result in discrimination, but not all cases of discrimination are also cases of stigmatization. It is the self-perpetuating, internally directed process that helps fully characterize stigmatization and contributes, along with externally enforced isolation, to poor health outcomes. As the legal scholar Scott Burris (2008) notes, a stigmatized person becomes "his own jailor, his own chorus of denunciation."

To help clarify this definition of stigmatization, it is useful to contrast stigma with a related concept, quarantine. Quarantine procedures, which date at least to the fourteenth century practice of isolating ships and passengers to prevent the importation of bubonic plague, also target undesirable traits to create physical segregation. Quarantine, however, primarily removes the trait from the social sphere through external enforcement, most commonly physical barriers and external social norms and laws. In contrast, while stigmatization can be achieved through external enforcement via physical barriers such as institutionalization, it is also maintained and enforced through internal mechanisms. Stigmatization works such that the stigmatized person also finds the trait undesirable. Shame about the marked trait encourages the stigmatized person to keep himself or herself apart from the broader community as opposed to quarantine, which is entirely externally enforced.

Ethical Dimensions of Stigmatization

Although the broad consensus among philosophers and social scientists is that stigmatization is almost always morally suspect, they have invoked several different ethical frameworks to reach this conclusion, most commonly versions of consequentialism or deontology. Consequentialist approaches can be broadly divided into two groups: act consequentialism, which focuses on the evaluation of a specific action, and rule consequentialism, which focuses on the evaluation of general rules or policies. In both cases, the key question involves whether a particular action or rule/policy has a net positive or net negative impact on well-being. Those that have a net negative result are morally impermissible.

While different versions of consequentialism define well-being differently – some focusing only on pleasure and others defining it as a constellation of desirable conditions related to human flourishing – they all focus on the net outcome of an act or policy as the appropriate focus of moral judgment. For an act consequentialist, we have to assess the impact of each individual act of stigmatization on the well-being of those involved (including the stigmatizer) to decide whether a given instance of stigmatization is justified. Because, however, most conversations about stigmatization are concerned with the global impact of the activity, epidemiologists and social scientists more commonly invoke rule consequentialism in considering whether we ought to allow stigmatization of a given trait or adopt policies that reduce the activity. Most rule consequentialists in the public health literature conclude that we are obligated to have policies that reduce or eliminate stigmatization because of its impact on health and the treatment and control of infectious and noninfectious stigmatized conditions.

In contrast to consequentialism, deontic approaches to the moral status of stigmatization focus on whether there is anything independently wrong with stigmatization, regardless of whether or not it has good or bad consequences for well-being. Like consequentialists, however, deontologists differ in their assessment of what might make an action or policy non-instrumentally

wrong or wrong on its own account. Some focus on whether stigmatization violates basic human rights or human dignity and others consider whether it is unjust or unfair for society to allow stigmatization because of the distribution or disproportional impact of the activity on one particular group. For example, Nussbaum argues the impact stigmatization has on social identity – narrowing the person to merely the stigmatized trait – is dehumanizing such that "we deny both the humanity we share with the person and the person's individuality" (Nussbaum 2006a). As such, it is always morally impermissible regardless of the net impact on well-being. Others have argued that stigmatizing actions or policies that allow or promote stigmatization are unjustified because they violate the basic human right to be treated with respect.

A third and less common approach to the moral status of stigmatization comes from virtue ethics, which focuses on the character traits a good person should develop in order to live well. These traits, in turn, inform the intentions, actions, emotions, values, attitudes, and sensibilities with which the good person approaches his or her interactions with others. On this account, because the impulses that lie behind stigmatization – fear, prejudice, discrimination, disgust, etc. – presumably do not fall as a mean between two virtues and are contrary to the reactions a good person would possess, we ought not to stigmatize. Philosophers and social scientists do not commonly invoke virtue ethics in discussing the moral status of stigmatization because the theory does not seem to capture what is wrong with stigmatization. A person should not merely avoid stigmatization because it is bad for his or her character or well-being but, more importantly, because of the impact stigmatization has on its target.

Importantly, consequentialists, deontologists, and virtue ethicists all assume that stigma is directly amenable to intervention. For example, Nussbaum writes extensively on the psychological mechanisms that drive stigmatizers and how the law and social policy might work to reorient this process. In contrast, empirical stigma researchers are often less optimistic that the forces that drive stigma can be easily averted. For example, psychiatrists Graham Thornicroft and Aliya Kassam (2008, p. 191) have argued that stigma research may not be actionable due to its focus on hypothetical rather than real situations and lack of "clear implications for how to intervene to reduce social rejection." While acknowledging that stigmatizing societal beliefs are indeed difficult to influence directly, there may be a different level at which to intervene, where appropriate. Strategies focusing on reduction of self-stigma have been found to successfully alter beliefs and enhance coping skills. Similarly, novel investigations exploring the complex interplay between poverty, illness, and stigma among HIV-positive women in sub-Saharan Africa have found that individual-level livelihood interventions may effectively reduce stigma by directly targeting poverty (Tsai et al. 2013). Ethicists should point to these efforts in identifying mechanisms through which unjustified stigmatization could be reduced.

Although consequentialists and deontologists reach the same conclusion in most cases of stigmatization, there is significant tension between the two over whether stigmatization is permissible in cases in which it appears to have overall positive consequences. For example, the sociologist Amitai Etzioni argues that stigmatization and shame have powerful deterrent consequences and that societies may be justified in harnessing this effect to prevent future criminal acts (Etzioni 2003). The most important public health example in the debate between consequentialist and deontologists regards social policies that stigmatize smokers to promote smoking cessation, although other examples include the stigmatization of unprotected sex as an AIDS prevention mechanism, the identification and shaming of sex offenders through registries and specialized license plates as a community protection mechanism, and proposals to stigmatize bullying to improve child and adolescent mental health.

In the case of smoking cessation, efforts over the last 30 years to socially isolate smokers and to use internal attitudes such as shame and guilt about smoking have clearly had an impact on overall smoking rates. For example, smokers in communities where smoking is rated as less acceptable are more likely to desire to quit, and these communities have overall lower smoking

rates and cigarette consumption. Changing attitudes about the social appeal of smoking has been a cornerstone of tobacco control policies. Consequently, part of the public health community's concern over the acceptance of electronic cigarettes in places in which tobacco smokers are excluded is that it will undo the positive effects of successful stigmatization of smoking behaviors.

Some public health rule consequentialists, pointing to the net overall impact on well-being of policies that encourage smoking stigmatization, argue that such policies are morally justified. Even though smoking stigma may add to poor health outcomes among individuals who continue to smoke, the benefit of reducing overall smoking rates justifies this consequence. In contrast, deontologists could argue that, insofar as these policies dehumanize smokers – for example, with advertisements that portray smokers as chimpanzees or equate smoking with pedophilia as in a series of public health posters in France – they are impermissible, regardless of the public health consequences (Burris 2008). They have also argued that, because smoking is more common and more entrenched in lower socioeconomic classes, stigmatizing polices are likely to unfairly impact already disadvantaged populations (Bell et al. 2010). Thus, even if the net result is a benefit to overall well-being, because the burdens are distributed unfairly, these policies are unjust and therefore impermissible.

One way to adjudicate this debate is to differentiate between policies that are frankly stigmatizing and those that aim to denormalize an activity. For example, the World Health Organization notes that denormalization aims to make tobacco use an undesirable practice by informing the public about smoking's negative consequences on health, society, the economy, and the environment (World Health Organization 2008). Here, the proposed mechanism through which smoking is made undesirable involves education and self-realization on the part of smokers rather than external prejudice involving negative stereotypes of smokers. Described this way, the denormalization of smoking does not clearly (or always) involve dehumanization or violations of basic human dignity. Or as Burris (2008, p. 475) puts it: "Fear of smoking, like the fear of syphilis. . . may contribute to stigma, but it is not itself stigma, and there is no reason not to promote it if we think it will reduce smoking rates."

The criminologist John Braithwaite similarly draws a distinction between shaming that is "reintegrative" – i.e., part of a process in which the relationship between the offender and the community is restored and the offender's identity repaired – from shaming that becomes stigmatization (Braithwaite 1989). Like Burris, he suggests that the use of disapproval and shame is ethically acceptable so long as it does not result in a level of rejection characteristic of stigmatization, in which there is no possibility of restoring a damaged social interdependence. Although such debates ultimately turn on how we distinguish between denormalization and stigmatization, it is important to emphasize that the broad consensus among philosophers and social scientists is that most cases of stigmatization are morally impermissible.

This section concludes with a short list of more complex questions about the ethics of stigmatization, discussions of which can be found in some of the works referenced in this entry. First, what steps are societies permitted to take to reduce or eliminate stigmatization? Is stigmatization itself permitted to stop stigmatizers? Second, does it matter ethically why humans stigmatize one another? Would it be more or less justifiable if stigmatization were an evolutionary instinct to protect from biologic contagions, a social instinct to protect communities from destabilizing influences, or a psychological defense mechanism that serves to reassert the stigmatizer's own strength when confronted with perceived deviancy? Third, what is the moral significance of the impact of stigmatization on the stigmatizers? Do separate obligations exist to reduce stigmatization because it is bad for the character of the stigmatizers?

Conclusion

Stigmatization is an activity that involves identifying and marking an undesirable trait, characteristic, or activity. It is enforced through external

social norms and policies and internally directed attitudes about having the marked trait including shame and disgust. Stigmatization has a substantial negative impact on public health outcomes, particularly for infectious disease and mental health, and often disproportionately affects already-vulnerable populations. Although stigmatization is widely held to be unethical, careful differentiation between stigmatization and denormalization may suggest that policies with a positive health impact such as smoking denormalization are justified.

Cross-References

- Access to Healthcare
- Behavior Modification
- Benefit and Harm
- Bioethics: and Politics
- Coercion
- Development and Bioethics
- Epidemiology
- Health Education and Promotion
- Health: Global
- Human Dignity
- Mental Illness and Institutionalization
- Obesity
- Paternalism
- Public Health
- Utilitarianism

References

Bell, K., Salmon, A., Bowers, M., Bell, J., & McCullough, L. (2010). Smoking, stigma and tobacco "denormalization": Further reflections on the use of stigma as a public health tool. *Social Science & Medicine, 70*(6), 795–799.

Braithwaite, J. (1989). *Crime, shame, and reintegration*. Cambridge: Cambridge University Press.

Burris, S. (2008). Stigma, ethics, and policy: A commentary on Bayer's "Stigma and the ethics of public health: Not can we but should we". *Social Science & Medicine, 67*(3), 474–475.

Corrigan, P. W., Larson, J. E., & Rüsch, N. (2009). Self-stigma and the "why try" effect: Impact on life goals and evidence-based practices. *World Psychiatry, 8*(2), 75–81.

Etzioni, A. (2003). *The monochrome society*. Princeton: Princeton University Press.

Goffman, E. (1963). *Stigma: Notes on the Management of Spoiled Identity*. New York: Simon & Schuster.

Hatzenbuehler, M. L., Phelan, J. C., & Link, B. G. (2013). Stigma as a fundamental cause of population health inequalities. *American Journal of Public Health, 103*(5), 813–821.

Link, B., & Phelan, J. (2001). Conceptualizing stigma. *Annual Review of Sociology, 27*, 363–385.

Marmot, M. (2006). Status syndrome: A challenge to medicine. *Journal of the American Medical Association, 295*(11), 1304–1307.

Nussbaum, M. (2006a). *Hiding from humanity: Disgust, shame, and the law*. Princeton: Princeton University Press.

Thornicroft, G., & Kassam, A. (2008). Public attitudes, stigma and discrimination against people with mental illness. In C. Morgan, K. McKenzie, & P. Fearon (Eds.), *Society and psychosis* (pp. 179–197). Cambridge: Cambridge University Press.

Tsai, A. C., Bangsberg, D. R., & Weiser, S. D. (2013). Harnessing poverty alleviation to reduce the stigma of HIV in Sub-Saharan Africa. *PLoS Medicine, 10*(11), e1001557.

World Health Organization. (2008). *WHO report on the global tobacco epidemic, 2008: The MPOWER package*. Geneva: The WHO Press.

Yang, L. H., Kleinman, A., Link, B. G., Phelan, J. C., Lee, S., & Good, B. (2007). Culture and stigma: Adding moral experience to stigma theory. *Social Science & Medicine, 64*(7), 1524–1535.

Further Readings

Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. New York: Simon & Schuster.

Heatherton, T. F., Kleck, R. E., Hebl, M. R., & Hull, J. G. (2000). *The social psychology of stigma*. New York: The Guildford Press.

Nussbaum, M. (2006b). *Hiding from humanity: Disgust, shame, and the law*. Princeton: Princeton University Press.

Strikes

Roberta Springer Loewy
General Medicine, Geriatrics and Bioethics, University of California, Davis, Gold River, CA, USA

Abstract

This entry addresses the phenomenon known as a "strike." It provides a brief introduction to the history and development of strikes in general.

Then, it describes the conceptual underpinnings and the sociological and ethical aspects of four different types of strikes: (1) general or "mass" strikes; (2) "capital" strikes; (3)"essential" strikes, which includes a section devoted to physician strikes; and (4) "hunger" strikes. This entry stresses the importance of understanding strikes and their related phenomena as specific incidences, each having its own unique biopsychosocial context, without which the process of ethical analysis cannot proceed. It discusses some of the ethical aspects of strikes in general and identifies those contextual features that are pertinent to assessing the ethical dimension in each type of strike. It concludes that (a) ethical analyses of strikes must be contextually thorough, (b) the ethical justification of particular strike actions will continue to undergo questioning and debate, and (c) while the right to strike is conditional because it must compete with various other conflicting rights, it is also necessary, for without it the ability to negotiate the boundaries of power relationships along the continuum between freedom and justice would be lost.

Keywords

Strikes; Collective actions; Protest actions; "Labor" strikes; "Capital" strikes; Justice; Equity; Civil disobedience; Power; Rights; Responsibilities; "Essential" services; Economic oppression; Interdependence

Introduction

This entry purposely attempts to avoid the professional language – and baggage – of professional ethics in order to describe, with as little obfuscation as possible, the problematic ethical nature of the phenomena of strikes. It eschews the basic traditional theoretic divisions into deontological (i.e., duty-based or principlist) and teleological (i.e., consequence-based) ethics as too unhelpful and misleading, especially within such an abbreviated format. Because such artificial divisions tend to privilege either duty (or principles) over consequences or consequences over duty (or principles), some of the important ethical elements of the problematic nature of strikes can easily be over- or underemphasized or even completely missed.

Thus, instead of prioritizing principles, duties, or consequences, this entry takes a more pragmatic, "all things considered" approach. Such an approach attempts to identify and weigh the value of such ethically laden concepts as duties, principles, and consequences, rights and responsibilities, and benefits and burdens as they unfold within the contextual relationships of specific strike situations. Moreover, it assumes not only that there is a fundamental interdependence between individuals but that this interdependence is a significant characteristic – not simply a defect – of persons. As a result the role that the benefits and burdens of all of those relevantly affected plays in ethical analysis is better captured and more fully appreciated.

History and Development: Background of the Issue

Conceptually, the idea behind a strike action has ancient roots, despite its modern name. In *The History of Trade Unionism* (1894), Beatrice and Sidney Webb claim such actions to have occurred nearly 1500 years before the Common Era. They cite the Biblical account of Exodus 5:7, wherein Hebrew brickmakers were denied their usual allotment of straw by a Pharaoh who ordered them to continue to make bricks with straw they themselves had to gather. Many treat the subsequent flight of these brickmakers from the city as the first record of a "strike" action. However, it is equally arguable that it was, in fact, the second recorded instance of a "strike" action – the first one (denial of straw) being the equivalent of a "capital strike" action by the Pharaoh.

According to Egerton (1951), the first historically documented account of what amounted to a successful strike action by workers occurred in 1152 BCE under Pharaoh Ramses III in Egypt when artisans of the Royal Necropolis at Deir

el-Medina ceased working because they had not been paid. Interestingly, here too, the focus has tended to be on the actions of the workers rather than the "capital" action of Ramses III (not paying his workers, which prompted their response).

In *The Outline of History* (1920), Wells notes that between the expulsion of the Roman kings (510 BCE) and the beginning of the first Punic War (264 BCE), plebeians sought and managed to win a greater share of governance through what we would recognize as several general strikes, twice actually marching out of Rome and threatening to establish a new city on the Tiber. And later (circa 100 BCE), the corruption of the powerful and pretentious collegia sodalicia (which were, essentially, trade associations who nominally adopted a deity so as to acquire greater economic and political clout) of Italy was met with protest actions equivalent to strikes – again, by an essentially politically abused and unrepresented populace.

It was not until the latter part of the eighteenth century that the word "strike" actually gained usage. According to the Webb and Webb (1894):

> The Oxford Dictionary gives the 1st instance of its use as in 1768, when the Annual Register refers to the hatters having "struck" for a rise in wages. The derivation appears to be from the sailors' term of "striking" the mast, thus bringing the movement to a stop.

In essence, the potential for strike actions (irrespective of what they might have been called in the past) has existed for as long as human interactions have been complex enough to develop power differentials perceived by some individuals' party to the interactions to be unduly coercive, unjust, or unethical. The more complex human associations become, the more opportunity for such power differentials to develop. Certainly, recourse to today's conception of a strike action has only increased since the beginning of the industrial revolution.

Conceptual Clarification/Definition

In its broadest, most general sense today, a strike is a collective form of protest toward an imposed condition or set of conditions generally felt by those striking as unduly coercive, unjust, or unfair. While most frequently associated with a collective refusal by those employed to work under conditions demanded by their employers, the word can just as readily appertain to collective refusals or demands by any given person or group of persons (public or private) and directed toward any other person or group of persons, institutions, or institutional policies (public or private, economic, political, or social).

Strikes arise for various reasons. These most commonly include disputes about wages and/or conditions of employment, jurisdictional disputes between unions, actions pursued for purely political goals (as in, e.g., a general strike or a hunger strike), or a combination of these (as in a "wildcat" strike, wherein a strike not authorized by the central body of a union may be directed by workers against any combination of employer, union leadership, and governmental institution or policies). While the underlying reasons that spark such protests may not be economic, the means almost always are. Hence, the vast majority of strike actions occur within – and affect or are affected by – the marketplace.

Less commonly appreciated, though of equal importance, is the distinction "capital" strike – capital being in quotes since today it is generally more narrowly construed to mean a corresponding action that can be taken by businesses. In this type of strike, those who own and/or control large amounts of wealth (financial capital) can, in effect, shut down the economy or manipulate governmental policy – whether by refusing to invest or reinvest in their infrastructure, by not hiring adequate workers, or by laying off existing workers.

The types of, reasons for, and frequencies of strikes, however, are never simply economic in nature; they are contingent upon a wide range of sociological factors including a country's history and general mores and the function/dysfunction of its social, political, and economic systems (including the role of its trade unions). These sociological factors can be quite long standing, subtle, and indirect.

While today capital tends to be narrowly defined as amassed wealth in the form of money

or other material assets, it actually encompasses a much broader understanding of the concept of capital (wealth). This broader conception of capital includes (though is not necessarily limited to) the following:

1. Natural capital – the kind of wealth associated with natural resources
2. Human capital – the inherited and acquired potential and capacities of unique individuals
3. Social capital – that fragile and elusive store of trust, mutual understanding, reciprocity, cooperation, shared values, and socially held knowledge which individuals share (or are denied)

Certainly not least of these is labor – for what else could be the source of all financial capital or wealth but an individual's human capital combined with natural capital (existing natural resources)? On this broader understanding of capital, the vast financial capital amassed by entrepreneurs today can be seen for what it is: derivative wealth or capital. While he might be accused of overstatement, Adam Smith declared in *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776):

> Labor. . .is alone the ultimate and real standard by which the value of all commodities can at all times and places be estimated and compared. It is their real price; money is their nominal price only.

When individuals or groups of individuals are deprived of access to the full range of opportunities (including social capital) normally available for maximizing their potentials or pursuing their interests and goals, they suffer injury: they are diminished as persons by being deprived of the respect due free and equal members of society and denied power they might – if given equal opportunity – otherwise have acquired. When individuals are denied control over their own labor (their own, unique capital or wealth), they not only suffer these same injuries; in addition, they will have been treated merely as a means to accomplish someone else's goals. Nine years after Smith's declaration, Immanuel Kant (1785) derived this same means/goals (ends) corollary in his practical formulation of the categorical imperative.

Hence, the right to strike is recognized, at least pro forma, by nearly all governments of the world. Some governments, however, may require a series of clearly specified, good faith efforts to arrive at a mutually agreed upon settlement prior to the strike; some may forbid purely political strikes or strikes by certain groups – the more common examples being public employees and "essential" workers, including healthcare workers, especially physicians and nurses.

While the purpose of most strikes or threats of strikes – whether by workers, businesses, or employer/owners – is to inflict some type of economic cost for failing to meet certain demands of those striking, some strikes are much more akin to demonstrations or political protests. These are often the result of a general class consciousness or occur in conjunction with an act of conscientious objection (see "▶ Conscientious Objection"). Therefore, when analyzing a strike's ethical dimension, it is essential to understand not only the goals (tacit, explicit, and those merely claimed pro forma) of those striking and the means utilized; it is also essential to understand the complex social preconditions – especially political and economic – that prompt recourse to a strike action.

Ethical Dimension

Strike actions demonstrate a rich and complex ethical dimension. They spring from environmental and social conditions; they are chosen by individuals; and they are most frequently acted out collectively. Moreover, they result in changes that have ripple effects on the whole: individuals, collectives, the physical environment, social conditions, and the society and its institutions and policies. In other words, they reflect the constant, ineradicable tension between individuals and their relationships to one another, to their physical and social environments, to institutions, and to the society as a whole. Depending on the structure of a society, that tension will weigh more or less heavily on the individual.

Strikes have been variously justified as "the lesser of two evils," as a moral right, as a legal

right, as an act of self-defense, as a basic right, and as a collective right. At their finest, strikes are attempts to reset and, thereby, stabilize the continuum between freedom and justice. This continuum is understood here as the communal efforts required to maximize the development and flourishing of individuals which, in turn, are necessary to secure the flourishing of community. In biopsychosocial terms, this resetting and stabilization could be thought of as a homeostatic biofeedback mechanism; in logical terms, it could be seen not as a tight, vicious circle but as a big, virtuous – albeit sometimes messy – one.

Unfortunately, even when goals and means are carefully articulated and organized, strike actions can still become quite destructive, certainly in the short term and often in the long term – as world history has amply and repeatedly shown. This is because strike actions are injurious to some extent or degree, and the act of injuring someone or something is never ethically neutral. Hence, depending on the nature and severity of the "injury" – which can range from minor inconvenience to major endangerment – it is open to the charge of being unethical unless compelling reasons can be marshaled that any alternative action or nonaction would be more injurious.

In 1945 the United Nations Educational, Scientific and Cultural Organization (UNESCO) came into existence with the aim of promoting peace and universal respect for justice, the rule of law, human rights, and fundamental freedoms. UNESCO's Universal Declaration on Bioethics and Human Rights (2005) identifies what are considered to be 15 principles (Articles 3–17) and their applications (Articles 18–28) that incorporate values that have been and continue to be widely recognized and shared by human beings across time and disparate cultures. Many of the values expressed within those articles – human dignity, personal integrity, personal responsibility, equality, justice, equity, solidarity, to name but a few – are the very values felt to be at risk by those who resort to strike actions. On this note, some of the ethical dimensions associated with specific kinds of strikes will be examined.

General Strikes

A general strike is a widespread collective action that can include political and social demands as well as purely economic ones. The strike force is usually composed of not only a wide swath of the workforce but, at times, students and members of the general citizenry as well. It is usually an action of last resort by a sector of the population who believe themselves or others to be disadvantaged by political and economic conditions. As a result, the goals of such strikes frequently are not primarily focused on addressing workplace demands from specific businesses or employers. Rather, they are more often focused on effecting political and/or economic public policy changes that, it is usually claimed, will better serve – and reflect – the public good. As such, they invite especial ethical scrutiny of the relationships between the citizenry, the workforce, and the government and its institutions (which may or may not include its economic ones).

Ethically speaking, the more democratic and open a government, the more a general strike action must be explained and justified as an action of last resort for protecting or restoring fundamental rights or interests. Hence, widespread or frequent general strikes within a democratic nation should be seen as ominous, representing significant challenges to the ethical legitimacy of the power relations and values underlying its very structure. When severe enough, such strikes may require a nation to undergo renegotiation of its basic social contract for it to remain a viable entity.

While strikes can and certainly do occur under any form of government, democracies that have not been well tended present a particularly challenging difficulty: an inherent power vacuum. Within such, the power of open and democratic processes can readily be subsumed by powerful, narrow interests. Unregulated capitalism is a most likely result. As Erich Loewy wrote in his text, *Freedom and Community* (1993):

> . . .capitalism is essentially and basically inimical to true democracy in that it seeks to place power in the hands of a relatively few who control the economic resources, whereas democracy seeks to place such control in the hands of the many. Democracy,

seeking to diffuse power, and capitalism, seeking to concentrate it, cannot well coexist.

Capitalism is basically an economic system. Democracy is basically a political system. Thus, capital need not – and, as a matter of fact, in today's environment, too often cannot – be contained within democratic border, making it difficult, indeed, for it to be adequately regulated within national borders or for any groups of individuals to mount an effective, nonviolent collective action against it. This merely serves to underscore the distinction between real and ideal, namely, that, in reality, democracies are no more immune to the unjust and inequitable concentrations of power into the hands of a few individuals than are any other forms of government. As Plato (circa 428–348 B.C.E.) reminds us in the *Republic*, it is immaterial what form of government a society has if the welfare of that society has been corrupted by greed, ignorance, incompetence, and/or overambitiousness.

So, irrespective of the form of government, capital (and its attendant power) is dependent upon labor since capital cannot come into existence without having been produced originally by someone's labor. When persons suffer because they are denied the just value of (and, therefore, power over) the fruits of their labor – whether by economic and/or political interests or means – they are ethically justified in seeking remedies. That being said, while there are often a range of remedies available, it is not only just but prudent to begin with the least disruptive remedies before resorting to those that are more disruptive so as to maximize both the good long- and short-term consequences while minimizing the bad.

"Capital" Strikes

A "capital" strike is initiated by those owning or controlling wealth (i.e., capital). As mentioned earlier, there are a number of ways wealth can be understood. "Capital" is used here in the narrower sense prevalent today: financial capital – any form of wealth that is available for use by entities who have amassed wealth to produce even more wealth. Financial capital can be owned, accumulated, and used by individuals, partnerships, or corporate entities and usually takes the form of material assets or cash.

A "capital" strike occurs whenever an entity or group of entities withholds capital either within a specific sector of economy or the economy as a whole. Such withholding may take several forms. It may be limited to individual businesses, wherein a company decides to pursue any one or combination of the following actions:

1. Not to replace retiring workers
2. Not to hire adequate numbers of workers
3. To lay off existing workers
4. Not to invest or reinvest in their business infrastructure

Or it may occur via a consortium of such entities who aim to stimulate changes in economic and/or political policy, whether at the local, state, national, or international level by, for instance, economic and/or political boycotts.

There are a number of conditions which may prompt a capital strike: companies may feel that their profit margins are unsustainable – or perhaps merely unacceptable – because of economic policies or government regulations. They may, for example, decide to sequester cash reserves rather than risk unfavorable loan terms. In short, when companies fear that returns on new investments of capital may, for whatever reasons, be unacceptable, inadequate, or nonexistent, a "capital" strike can result. In such cases, the resulting strike is basically reactive.

But "capital" strikes can be proactive as well: wealthy companies or entities can exert a powerful influence over the development of economic policies and public and private regulatory agencies. Such power – and the fear of its loss – too often results in an ethical distortion of interests, values, and goals. To the degree that a company or entity pursues narrow self-interest to the exclusion of considering the basic interests, values, and goals of those with whom it interacts – be they other businesses, workers, consumers, or the rest of society – it causes disproportionate and unjustifiable harms by benefitting itself at the expense of others. This calls into question not only the ethicality of the company's or entity's

present behavior but ultimately its long-term viability as well. For, after all, the well-being of a society depends on the well-being of its individual members which, in turn, requires the society to ensure a reasonably equitable and justified distribution of both benefits and harms.

"Essential" Strikes

"Essential" strikes are collective actions taken by individuals who work in areas – public or private – that can be categorized as essential to the provision of basic human needs. Essential services can include, but are not necessarily limited to, employment in the following areas of service: police, military, public health, medical (including medical facilities and pharmaceutical industries, etc.), emergency response teams, water, energy, postal, and communications related (reportage, journalism, television, radio, the Internet, telephone, etc.). Because there is often only one communal source for many of these essential services, strike actions of this sort can be more than simply inconvenient or annoying – they can be quite harmful and disruptive to the functioning of individuals as well as of the community as a whole.

Some have argued that workers in these areas should have no right to strike at all because of the potential dire consequences that could result from disruptions in the provision of these essential services. Others have argued that to deny the right to strike to anyone – "essential" workers included – would be tantamount to enslavement because loss of the right to strike would entail loss of recourse to effective collective bargaining (since collective bargaining without leverage from each side to the dispute would rob the process of the balance necessary for the most equitable resolution).

Given these two poles of argument, most governments recognize, at least in principle, a constitutional right to strike by all workers, including "essential" workers while reserving the power to subject that right to certain limitations and/or conditions. The rationale behind such regulation is based on the claim that, for any right to strike to remain ethically tenable, it must always compete and be balanced against other fundamental rights: in the case of an "essential" strike, it must compete and be balanced against the basic needs required for both the society and its individuals to flourish. As discussed earlier in this section – and echoed in Article 14 of UNESCO's Declaration (2005) – the promotion of the best interests of all persons (social development, access to basic needs, reduction of poverty, illiteracy, exclusion and marginalization of any individuals, etc.) is "the central purpose of governments that all sectors of society share."

While strike actions by police and military services stand, rather obviously, as the most stringently regulated by public policy, any strike by "essential" workers may, at some point, be regulated to the extent that it is reasonable and can be justified as the least burdensome – yet still effective – ethical alternative available. One of the ways such balance might be achieved in lieu of outright strike action is through mandatory (binding, compulsory) arbitration, where, unlike collective bargaining, all parties to a dispute regarding essential services are required to submit to a neutral third-party arbitration process.

As a result, most essential workers are expected to submit either to collective bargaining or mandatory arbitration in lieu of strike action. Yet, tellingly, even this policy has not completely resolved the issue. For example, in the USA police officers have been known to resort to what has euphemistically been called "the blue flu" (a quasi-strike action known as a "slowdown") when arbitration becomes deadlocked or the officers believe their position has not been adequately or fairly addressed.

Interestingly, potential strike actions by healthcare professionals – especially physicians – have, at least until recently, been viewed somewhat differently, viz., few governments have been eager to resort to mandatory arbitration or legislative regulation specific to strike action; rather, they have been content to defer to the professions' own autonomously endorsed ethical precepts to prevent such action. Hence, it is instructive to examine why this still tends to be the case. Since much of our familiarity with professional ethics comes as a result of the rich scholarship that has occurred in the field of medical ethics during the

last 70 years or so, the following section will deal specifically with physician strikes.

Professionalism

Most of the workers involved with essential services are broadly considered "professionals," even when they do not meet all of the traditional characteristics of a profession, viz., a group of individuals with exclusive service-oriented expertise and a specialized body of knowledge over which they are granted control (education, apprenticeship, licensure, regulatory power to admit and discipline members, etc.). Society not unreasonably expects a proportionate return for bestowing professionals with such considerable power over their respective disciplines, holding professionals to a heightened degree of responsibility equivalent to the heightened degree of professional autonomy granted them.

While their education is usually heavily subsidized, traditional professions are largely self-governing, insofar as they control the structure, education, and standards of practice and the licensing, regulation, and censure of their members. Physicians are one of the very few groups who belong to a profession in this more narrow, traditional sense of the term. In return for this high level of autonomous expertise, physicians undertake a fiduciary obligation to society: a trust relationship with patients – to care for them and to avoid doing them gratuitous harm. This includes assuming indirect as well as direct responsibility for coordinating and overseeing care (even when not present), never abandoning patients and never providing minimal or suboptimal care.

Under such a traditional fiduciary relationship, unless patients can be carefully safeguarded, physicians may never strike for purely self-interested reasons. Even when a strike can be justified in terms of being in the best interests of patients generally, physicians cannot knowingly place patients at risk. However, this traditional fiduciary relationship has undergone severe stress with the rise of third-party nonmedical interests (insurance companies, healthcare organizations, regulatory agencies, etc.) bent on acquiring increasing control over the costs – and, thus, the practice – of medicine today. As a result, patients and physicians are being constrained in novel ways, and the traditional taboos regarding physician strikes are beginning to erode.

Professionalism, Trust, and Its Erosion

According to Thomasma and Hurley (1988), there are three basic causes for physicians to strike:

1. To bring about better patient care for present and/or future patients
2. To contest third-party intrusions
3. To obtain better pay or benefits

As with other forms of strike, much depends on the existing social and political conditions and stresses: certainly, under circumstances where facilities, available therapies, or standards of care are shoddy or lacking, physicians are more likely to feel a strike action justified in order to protect current and future patients. Likewise, when third-party intrusions threaten patient safety – especially when they thwart or gain control of governmental regulatory powers – physicians are more likely to see some form of strike action as an option. And finally, if large numbers of physicians become unable to provide for their own basic needs due to a lack of control over the conditions of their practice, they are much more likely today to entertain strike actions – and are also more likely to justify them, at least in part, by claiming that the existing state of affairs is a threat to patient safety.

Over 35 years ago Norm Daniels argued in *On the Picket Line* (1978) that there are four criteria for a strike action by healthcare professionals to be ethical:

1. The strike action cannot result in serious loss of life.
2. The goals of the strike action must include improvements in patient care.
3. The pursuit of good faith efforts to resolve the issue must first have been attempted prior to any strike action.
4. The target and goals of the strike action must be carefully articulated.

Over the intervening years these criteria have been endorsed – at least in theory, if not so much

in practice – by physicians as well as other healthcare workers. However, it might be argued that this list could be strengthened by the addition of several more criteria:

1. The identification and utilization of the most humane organization possible of the ***means*** by which the goals of the strike action are to be pursued.
2. Any proposed strike action should pursue a "least harm" policy:
    1. The least burdensome, while yet effective, type of strike action should be chosen – e.g., choosing a "paper" strike, where medical care is given but charting is stopped for all but seriously ill and unstable patients.
    2. Strikes must always be weighed against the harms of not striking.
3. Widespread notification and public discussion of any contemplated strike action – this, alone, may stimulate negotiation sufficiently to eliminate the need to strike.
4. Because of the social investment made in the education and training of physicians, any agreements reached for settling physician strike actions should be subject to public scrutiny, discussion, and approval.

While the pursuit of such strategies will not guarantee a halt to the erosion of professional autonomy or the contingent loss of trust increasingly felt by patients and professionals today, it may slow their progress until a healthcare system with a more humane, homeostatic balance is adopted.

Hunger Strikes

A hunger strike is the nonviolent protest action of refusing to eat by a person or group of persons. Hunger strikes are undertaken to call attention to grave issues – objectionable conditions, societal injustices, imprisonment, improper treatment, etc. – and to force change. Hunger strikes differ from the traditional understanding of a strike, insofar as the effects of the strike action are most immediately and profoundly felt by the striker or strikers and only indirectly by those against whom the strike is called. Instead of threatening economic hardship, hunger strikers rely on their own suffering (see "► Suffering") to raise awareness and induce feelings of shame, guilt, and sympathy not only in those against whom they are striking but also in those who are bystanders to the situation. Once again, the success of such actions depends upon the specific context – the social and political conditions – within which such actions occur.

Unlike most other types of strikes, many of the most memorable hunger strikes have been accomplished by individuals – the twentieth century British suffragette Marion Wallace-Dunlop; the Indian leader Mohandas Gandhi; the Latin-American civil rights leader Cesar Chavez; the Cuban journalist and dissident Guillermo Farinas; the Nigerian poet, playwright, and Nobel laureate Wole Soyinka; and the Irish nationalist Bobby Sands are a sample of well-known individuals.

However, there have been notable mass hunger strikes as well. In 1922–1923, for example, more than 14,000 Irish republicans were arrested without charge and kept in prison and internment camps without trial and endured significant deprivation and suffering. They struck for 41 days. In 2013, 30,000 of California prison inmates, living in some of the most oppressive prison conditions possible, began one of the largest hunger strikes in the USA to protest the use of long-term solitary confinement. One of the largest and longest mass hunger strikes was the rolling hunger strike (short-term, relay-style fasts) begun in 2006 by hundreds of Chinese activists. After the Chinese government responded with beatings, house arrests, kidnapping, and jail, tens of thousands of supporters in the international community joined in solidarity.

Perhaps the most extraordinary hunger strikes of recent history (circa 2002–2013) occurred at the US detention camps at Guantanamo Bay in Cuba. Prisoners, most of whom were being held without having been charged, struck in hopes of being treated in a manner consistent with the Geneva Conventions. The USA responded by force-feeding the prisoners with the help of medical personnel (most assuredly a prime example of medical professionals "treating" a government as the "patient" rather than the striking prisoners).

Despite being pressured by the United Nations, the USA ultimately buried the situation by simply announcing that it would no longer release information about the hunger strikes as such information "serve[s] no operational purpose" (The Huffington Post 2013).

That individuals feel required to subject themselves to such self-inflicted harms should generate concern for both the hunger strikers and the social conditions – perceived or real – that prompted such action. After all, the dignity and integrity of a society are measured largely by the importance it attaches to protecting the dignity and integrity of its members – especially of those deemed most vulnerable. As reflected in Articles 3 and 8 of UNESCO's Declaration on Human Rights and Bioethics, dignity and integrity are foundational concepts for the preservation not only of autonomy but of justice as well.

Conclusion

Strikes, as can be seen from this brief entry, are various and ethically complex. They arise in response to unjust or unfair power differentials, whether perceived or real. They may occur between individuals, between individuals and corporate entities, between individuals and governments, or any combination thereof. Whatever their origin, they are significant insofar as they represent a serious, concrete objection to and rejection of existing practices or policies. They are predominantly economic and/or political in nature and are usually considered a stimulus for discussion, negotiation, and reform and, in the final analysis, they remain – and ought to remain, according to most assessments, pro or con – a measure of last resort.

Because of the variety of conflicting values, interests, and goals of the participants involved, ethical analyses of strike actions must be contextually thorough. Because of the unique circumstances and rich context surrounding each individual case, the ethical justification of strike actions in general will continue to undergo questioning and debate. And, finally, the right to strike must be considered a conditional right because it is always in competition with various other rights. However, it must also be considered a necessary right since, absent the right to strike, there is no effective leverage against coercive, unfair, or unjust institutionalized power – be it public or private, be it economic and/or political.

Cross-References

- ▶ Applied Ethics
- ▶ Autonomy
- ▶ Biopolitics
- ▶ Civil Disobedience
- ▶ Coercion
- ▶ Commercialism and Healthcare
- ▶ Common Good
- ▶ Compassion
- ▶ Conscientious Objection
- ▶ Egalitarianism
- ▶ Equality and Equity
- ▶ Exploitation
- ▶ Governance
- ▶ Human Dignity
- ▶ Human Rights
- ▶ Justice: Global
- ▶ Justice: Theories of
- ▶ Mediation
- ▶ Professionalism
- ▶ Professional-Patient Relationship
- ▶ Social Ethics
- ▶ Solidarity
- ▶ Suffering
- ▶ Trust
- ▶ Vulnerability

References

Daniels, N. (1978). On the picket line. *The Hastings Center Report, 8*(1), 24–29.

Egerton, W. F. (1951). The strikes in Ramses III's twenty-ninth year. *Journal of Near Eastern Studies, 10*, 137–145.

Kant, I. (1785 [1958]). *Foundations of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis: Bobbs-Merrill Educational Company.

Loewy, E. H. (1993). *Freedom and community: The ethics of interdependence*. Albany: State University of New York Publishers.

Smith, A. (1776 [1937]). *An inquiry into the nature and causes of the wealth of nations*. New York: Modern Library.
The Huffington Post. 4 December 2013.
Thomasma, D. C., & Hurley, R. M. (1988). Ethics of health professional strikes. In D. C. Thomasma & J. F. Monagle (Eds.), *Medical ethics: A guide for health professionals* (pp. 358–374). Rockville: Aspen Publishers, chapter 31.
UNESCO. (2005). Universal declaration on bioethics and human rights. http://portal.unesco.org/en/ev.php-URL_ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html. Accessed 1 Oct 2015.
Webb, B., & Webb, S. (1894 [1976]). *The history of trade unionism*. New York: AMS Press.
Wells, H. G. (1920 [1931]). *The new and revised outline of history*. New York: Garden City Publishing Company.

Further Readings

Picketty, T. (2013). *Capital in the twenty-first century* (A. Goldhammer, Trans.). First published as Le capital au XXI siecle. The Belknap Press of Harvard University Press: Cambridge (MA) and London (England), 2014.
Pike, J. (1998). Strikes. In R. Chadwick (Ed.), *Encyclopedia of applied ethics* (2nd ed., Vol. 4, pp. 239–247). San Diego: Academic.
Stiglitz, J. (2012). *The price of inequality: How today's divided society endangers our future*. New York: Norton and Company.

Substance Abuse

Ambros Uchtenhagen
Swiss Research Institute for Public Health and Addiction, Zurich University, Zurich, Switzerland

Abstract

Psychotropic substances were used throughout history for medicinal, ritual, and recreational purposes. Overuse may result in pathological consumption styles, mainly dependence and other forms of abuse (hazardous use, harmful use). Attempts at political, religious, and social control of use, including total prohibition of all or specific substances, show mixed success. While traditional use and experimental use mostly have a sociocultural background, dependence and other forms of abuse are classified as a behavioral pathology and as medical conditions. Different ethical aspects and rules apply for use, nondependent abuse, and dependence, respectively. Human rights and medical ethics provide a framework of consequential ethics. Interventions for abuse, their organization, and performance are discussed in some detail, and special attention is paid to the vulnerabilities (stigmatization, marginalization, comorbidities) of people with addictive behavior.

Keywords

Drug abuse; Addiction; Medical ethics; Public health; Drug policy; Service research

Introduction

Substance abuse is one of the major health and social problems at global level, in spite of all efforts to curb extent and ensuing damage. Alcohol and tobacco rank highest of all psychotropic substances as contributors to the global burden of disease (Murray and Lopez 1997). An estimated 166–324 million people have used an illicit drug in 2012, and 18–39 million are abusers (UNODC 2015).

The relevant UN Organizations recognize substance abuse as a medical condition for which evidence-based preventive and therapeutic interventions exist. The implementation of the respective services and professional education is far from satisfactory. Only one of six dependent users in need of treatment receives it; even more alarming is the low access of imprisoned drug users to appropriate treatment (UNODC 2015). The principles and rules of medical ethics are often overruled by other interests, moral prejudice, and discrimination. Enormous efforts at all levels are at stake for improving effective and ethically acceptable care for substance abuse.

Substance Use and Abuse History

Use came before abuse. Psychotropic substances have been used throughout the history of

mankind, starting with the knowledge of plants and their effects in the age of hunters and gatherers. The observation of fermented fruit and the production of alcoholic beverages started well before the 5th Millennium B.C., for nutrition and ritual purposes. Among the first documented in history are also opiates, detected in human dwellings from the 3rd Millennium B.C. Alcohol, opium, cannabis, coca, and tobacco have a long history of medicinal use, as universal or especially for analgesia. Other therapeutically/ritually used drugs are hallucinogenic substances, khat and kava.

Problematic effects of use must have been observed early on, and we have relevant documents on such observations since antiquity. Problematic effects may explain the development of consumption control strategies, including privileged access (e.g., coca for the ruling class in Inca culture, hallucinogenic drugs for initiates only in Greek secret societies, wine for nobility over centuries, rationing schemes, use of drugs at specific events only). Ritualistic arrangements allowed use without excessive use. Main control strategies came from political and religious authorities. The notion of substance use and misuse as a moral weakness grew on the basis of a religious ideal of moderation or abstinence from pleasure-seeking behavior. It was a reaction to abundant substance use and its negative consequences. In medieval Europe, restrictive alcohol use was imposed after the deadly epidemic of plague. Alcohol and drugs are banished in Islam; alcohol prohibition existed from 1919 to1933 in the USA. Rationing schemes are known for opium in Middle East countries, for alcohol in Sweden. Other strategies were implemented by civil societies (temperance movements) since the eighteenth century, when the mass production and consumption of spirits created major social problems across all strata of society and threatened the increasing economic competition and the demand for efficient management. Finally, the extraction of active substances from the natural products – morphine and cocaine – and their use in patent medicines in late nineteenth century, as well as the invention of parenteral injection, led to an enormous increase in consumption. This prepared the ground for legal measures in order to curb the extent of abuse, on national and international levels (Harrison Act in 1914 in the USA, International Opium Convention in 1912, UN conventions on narcotic drugs in 1961, 1971, 1988).

All these efforts to control use and abuse of addictive substances did not primarily focus on the individual, but on society or on specific strata of society, in the interest of public health and public order. The concepts of "abuse," dependence, and addiction however developed in the medical field, based on observations of patients suffering from negative consequences of their habit. Alcoholism was described as a medical condition in the course of the nineteenth century, followed by morphinism and cocainism. Various terms were used for a similar clinical syndrome, while the concept oscillated between a physiological illness, a mental condition, and a learned misbehavior. Therapeutic regimes of addicted patients evolved. An abstinence regime was introduced and recommended when the condition was attributed to the effects of the substance, without regard to environmental and personality factors. Psychotherapy, milieu therapy, and spiritual interventions grew on the respective beliefs about the relevant cause of addiction.

Conceptual Clarification/Definition

Current Diagnostic Criteria

The medical concepts of substance abuse and dependence define those as disorders, comparable to other medical conditions, with specific symptoms; with somatic, psychological, and social risks; and with evaluated therapeutic approaches. This is in contrast to other interpretations of addictive behavior, mainly a moral understanding (weakness of will, undisciplined pleasure seeking, egotistic neglect of social obligations, etc.). Such moralistic attitudes are widespread, stigmatizing addicted persons and impeding their treatment. Another complication comes from legal prohibition, when substance use and abuse are considered criminal acts that must be punished.

Substance Abuse, Table 1 Diagnostic criteria for substance-related disorders

DSM-5	ICD-10
Addiction and related disorders Substance-specific disorders (12 types) *Criteria for a substance-related disorder* (11) *Failure* to fulfill major role obligations *Recurrent* physically hazardous use (e.g., driving while intoxicated) *Recurrent* or persistent social/interpersonal problems due to alcohol *Tolerance* *Withdrawal* *Use more* or more often than intended *Persistent* desire to cut down or control use *Too much time* for purchase, use, and recovering *Important* activities neglected *Continued* use in spite of knowing negative health effects *Craving* *Severity specifier* *Severe*: 4 or more criteria positive in a 12-month period *Mild*: 2–3 criteria positive *No disorder*: <2 criteria positive Specify if with or without physiological dependence *Course specifiers* *Early* full remission, early partial remission *Sustained* full remission, sustained partial remission *On agonist* therapy *In a controlled environment*	**Mental and behavioral disorders due to psychoactive substance use** *Harmful use* A pattern of psychoactive substance use that is causing damage to health. The damage may be physical (as in cases of hepatitis from the self-administration of injected psychoactive substances) or mental (e.g., episodes of depressive disorder secondary to heavy consumption of alcohol) *Dependence syndrome* A cluster of behavioral, cognitive, and physiological phenomena that develop after repeated substance use and that typically include a strong desire to take the drug, difficulties in controlling its use, persisting in its use despite harmful consequences, a higher priority given to drug use than to other activities and obligations, increased tolerance, and sometimes a physical withdrawal state The dependence syndrome may be present for a specific psychoactive substance (e.g., tobacco, alcohol, or diazepam), for a class of substances (e.g., opioid drugs), or for a wider range of pharmacologically different psychoactive substances

The current diagnostic definitions are part of the International Classification of Diseases by World Health Organization, 10th edition (WHO 1990), and of the Diagnostic Statistical Manual of the American Psychiatric Association, 5th edition (APA 2013). Those definitions use different terms and specifications (Table 1). This limits the comparability in substance abuse research, as well as the communication on clinical and public health issues in substance abuse.

The criteria in both diagnostic systems are heterogeneous: some physiological (tolerance, withdrawal), some psychological (craving, desire to stop or reduce consumption, loss of control), or social (neglecting other activities). For determining a diagnosis, only some of the criteria are to be met; consequently, patients with the same diagnosis present heterogeneous clusters of criteria. Criteria are valid for all types of substances, but have not the same weight for all substances. Finally, the various addictive substances differ widely in their risk profiles, and therefore the term substance abuse covers a range of substance-specific issues.

Many of these criteria apply also to other human behavior. Withdrawal, neglect of other activities, and continuation despite problematic consequences may apply to "workaholics" or excessive sexual appetite, but we value it quite different from our value judgment on substance abuse and dependence.

In the following, the term of substance abuse has the meaning of any kind of hazardous, harmful, or pathological substance consumption, substance dependence being a specified type of abuse. Problem drug use is about the same as abuse; addiction is a synonym of dependence.

Addiction Theory

The diagnostic systems are essentially descriptive, avoiding causal factors for the definition of disorders. They do not explain how and why addictive behavior happens. In contrast, the present understanding of substance abuse and addictive behavior

in general (including syndromes of nonchemical dependence) focuses on brain research, epidemiology and clinical studies on personality factors, prognosis, and treatment outcome.

Based on the development of brain imaging techniques in vivo, an enormous amount of knowledge accumulated on functional processes as effects of substances of abuse, as well as on the structure of the cerebral reward system and its transmitters, and on structural changes after continued use of such substances (Volkow et al. 2003).

Epidemiological surveys provide knowledge on incidence and prevalence rates of substance use and related disorders, on how substances are used, by whom, and with what kind of consequences; repeated surveys can tell us about changes over time. Knowledge on the social conditions that shape addictive behavior and on the role of interventions at population level stems from this type of research (WHO 2002).

Patient observation and clinical research provides knowledge on personality factors and psychological processes involved in addictive behavior, as well as on the value and limitations of therapeutic interventions at individual level.

Neither of these research-based theories is sufficient to explain addiction. Theory of addiction at its best makes an effort to bring these different types of knowledge together and to consider their interactions and complimentary value rather than opposing them. A prominent example is the volume by Robert West (West 2006), presenting a comprehensive research-based framework of motivational processes as a basis for human behavior. In general, there is some consensus in the scientific community about the diversity of factors – somatic, psychological, environmental – contributing to the development of substance abuse. The most popular underlying concept is the biopsychosocial model of George Engel that stresses the need for an interdisciplinary approach in dealing with most psychiatric conditions including addictive disorders.

Different types of explanation are valuable for specific phases in the development of addiction. Starting the use of addictive substances is mainly fostered by milieu factors, social conditions, peer influence, self-regulation/self-medication to counteract unpleasant states of mind, and certain personality traits such as sensation seeking. The passage from occasional to regular use, including hazardous and harmful use, is facilitated by a trend to optimize performance (neuroenhancement), by successful self-regulation, but also by a deficit in other forms of pleasurable experiences. Developing dependence finally is facilitated by genetic vulnerability and exposure to chronic stress (internal or external).

The very nature of dependence is far from being universally accepted. On the one hand, dependence is considered a chronic, relapsing condition; on the other side, the majority of persons with substance dependence recover without formal or therapeutic interventions. The lessons to be learned from "natural recovery" include the role of life events and of social support for successful and persistent behavior change from dependent use to moderate use or abstinence.

Ethical Dimension

Ethical aspects of substance abuse cover a range of issues: the ethics of substance use as well as of abuse, in an individual and in a societal perspective. They need a separate discussion. The common framework is the concept of consequential ethics.

Basic Orientation: Consequential Ethics

The philosophical debate includes a distinction between absolutist and utilitarian positions. Absolutism means the acceptance of a conduct code based on absolute, indisputable rights and duties (e.g., abstaining from substance use). Utilitarianism has its focus on the consequences, not on the reasons or motives of conduct; whatever the motives are, moral judgment is based on the consequences of behavior (e.g., of substance use). This utility principle has a long-standing tradition in various forms. A most prominent representative was the English philosopher John Stuart Mill; a contemporary representative of a consequential ethics is Hans Jonas.

The absolutist position states: "Right is to be done come what will come. I am not answerable for the consequences of doing right, only of not doing it." The utilitarian position is summed up by "success is the touchstone; the might of obtaining the reward." Examples of absolutist positions are the categorical demand for abstinence from all or specific addictive substances or the claim for complete individual responsibility in using such substances without interference by others. A utilitarian position however cares about the consequences of use as well as of any interventions against substance use and abuse, in general and under specific sociocultural conditions.

Ethical Aspects of Substance Use

Acceptability of substance use depends on the acceptability of the consequences of use. Consequential ethics must value the outcome of substance use for individuals and for society.

The risk profile of specific substances should be known and respected. The Global Burden of Disease Study of World Health Organization documents the prominent role of legal substances (alcohol and tobacco) for health problems at the population level, and a recent summary describes the extent of addictive behaviors globally (Gowing et al. 2015). Ratings of substances according to their harm profile are available (Van Amsterdam et al. 2015).

The harm profile has not consistently guided drug policies and the ethical debate. Alcohol, in spite of its high-ranking position concerning harms, has received some attention in terms of harm prevention, but public opinion and policies are reluctant to envisage effective strategies to curb consumption. This is due to long-standing cultural traditions, to economic considerations, and to the influence of a powerful industry profiting from extensive consumption. On the other side, the prohibition of substances with a much lower rating for harm (e.g., cannabis or ecstasy) does not respond to a consistent policy orientation of harm avoidance, but to other political considerations. These are examples of ethical conflict, weighing the potential and effective harm of a specific substance against other societal values. A cultural embedding of use in a long-standing tradition favors acceptability of use in spite of major harm, while foreign origin of "newcomers" in the drug spectrum favors repression.

Apart from these culture-specific aspects, the ethical debate also knows some global arguments concerning individual and societal values. At the individual level, the main values at stake are the demand for self-fulfillment (developing the personal talents and potentials), self-responsibility (managing one's own life with sensible goals and decisions), and self-control (limitation of pleasure-seeking behavior). At the societal level, the values at stake are the fulfillment of citizen's obligations (for one's own sustenance and for the functioning of the community), maintenance of public safety and public order, and the sociocultural acceptability of behavior. For all these, the use of addictive substances may have consequences, and acceptability is linked to the extent of negative effects of use.

Ethical Aspects of Nondependent Substance Abuse

The ethical aspects of nondependent abuse (hazardous use, high-risk use, harmful use without meeting the criteria for dependence) must be discussed separately from those of dependent use. Nondependent abuse has its roots in sociocultural conditions, in learned behavior, in the characteristics of a given professional or family milieu, and in a person's attitudes, and it is not generally understood to be a medical condition in the same sense as dependence. Nevertheless, nondependent abuse is included in the diagnostic schemes (per se or as a less severe form of abuse) and is frequently diagnosed in medical practice. Specific interventions (brief interventions, early interventions, motivational interviewing) are available and applicable wherever nondependent abuse is seen (in medical and social services, by police, in clubs, etc.). They are evaluated to be effective. The professional ethics of those dealing with nondependent abusers apply; there are no universal rules to observe.

However, nondependent abuse precedes dependent abuse, and hence, interventions against the social factors fostering substance abuse are an essential instrument to reduce dependence.

According to epidemiological research, economic inequality, illiteracy, misguided urbanization, and loss of social traditions and networks are important causal factors for substance abuse globally (WHO 2002). In order to achieve the WHO goal "health for all," the scope of activities must include interventions securing the conditions for health protection. Great efforts are necessary on all levels, national and international; in legislation; in health policy; in the education of all concerned professionals; and in service provision in order to reduce incidence and prevalence rates of drug abuse.

Ethical Aspects of Substance Dependence

Which Ethics Apply?

Today's drug policy claims to be evidence based. Evidence means that policy recommendations follow scientific findings on "what works" and therefore have a good chance to lead to positive results. Drug policy is based on principles of consequential ethics. But what are the goals? We have to deal here with the criteria of the human rights declaration as a general framework and with the principles of medical ethics that must fully apply when dealing with substance abuse. Will they allow us an ethical judgment on the results of treatment as well as on the consequences of treatment policy?

Human rights: The principles of human rights apply to all human beings. Most states have signed the Universal Declaration of Human Rights (UN 1948). The Universal Declaration promotes a number of relevant conditions:

- No discrimination (art.2).
- No degrading or inhuman treatment (art.5).
- Right of equal access to medical care and social services (art.25/1).
- Everyone has duties to the community (art.29/1).
- Limitation of rights and freedom are admissible on the basis of "just requirements of morality, public order and the general welfare" (art.29/2).

The European Convention on Human Rights (Council of Europe 1950) further stipulates that a person's liberty may be deprived in case of lawful detention of alcoholics or drug addicts (art.5/1/e), with a right to appeal to a court (art.5/4).

These statements try to establish a balance between protecting the individual rights of the person and respecting the needs of society for public order, general welfare, and even morality. There is large room for interpretation, so that every society can decide on the compatibility of addictive behavior with the nature and extent of these requirements. Compulsory measures against persons with substance dependence may be admissible by national law. Therefore, human rights are no universally accepted basis for dealing with substance dependence.

Medical ethics. More to the point are medical ethics, as far as substance dependence and all types of substance-related conditions are accepted to be medical conditions. Substance dependent persons are patients and should enjoy the status and rights of patients in general. Principles of medical ethics apply in all matters of diagnosis, treatment, research, staff attitudes, service provision, and health policy.

However relevant the medical ethics are, they are not sufficient to cover all aspects of substance dependence. This is the basis of two controversies: the so-called medicalization of a social problem and the fight of medical versus moral treatment. Treating patients without regard to the social factors (at individual and population level) that facilitate incidence and prevalence of medical conditions is a misconception; Public health is a concept based on the necessity to care for the social conditions of illnesses. The basic ambivalence – substance dependence as a medical condition or a moral weakness – is reflected in the opposition of medical and moral treatment. If moral treatment is understood to be educational and admonishing, this approach is nowadays widely replaced by an approach to help the addict in getting motivated for behavior change, by appropriate empathy, information, social, and moral support.

Medical Ethics: General Rules

The general ethical rules for good medical practice apply. The so-called oath of Hippocrates

stated already to respect the patient's well-being as the highest value in medicine and to respect the principle of the medical secret perpetually. A modern form of doctor's obligation is the "Declaration of Geneva" of 1948 by the World Medical Association. It states that "the health of my patient will be my first consideration."

Other sources on medical ethics are national conventions; examples are the Standards of Conduct of the American Medical Association and the Good Medical Practice of the English General Medical Council. Main issues are the patient's autonomy of decision, informed consent, dignity and confidentiality, nondiscriminatory beneficence, and nonmaleficence, but also keeping up professional standards by continued education and networking with other services and colleagues in order to provide the best possible care. The American recommendations also include a responsibility to seek change in official or legal requirements that are contrary to the best interests of the patient. These codes usually acknowledge the occurrence of ethical conflicts and provide links for support in such cases.

Specific Ethical Rules for Substance Abuse Patients

It is obvious that even these few principles cannot be followed without creating conflict. Imminent risks from intoxication and risks of chronic self-damage invite measures to avoid causing harm to the abuser and to improving his/her well-being. What if the abuser does not comply? Involuntary intervention to prevent harm for the abuser as well as for other persons is in conflict with the autonomy of an unwilling patient. Treating all abusers as being equal is difficult when following the lessons from evidence-based guidance for interventions. The protection of sensitive data in the interest of substance abusers is often in conflict with administrative and law enforcement interests in case of illicit drug use or risks for others.

Such conflicts must be carefully examined, in the best interest of all concerned. Some rules on how to deal with conflicts apply. When the patient's interests collide with those of relatives or other third parties, mediation must take place for a common solution. It is advisable to recur to an ethical *consilium* (second opinion from an expert) if major consequences are expected from a contested decision. In general, the principle of respecting the autonomy of the patient must never be overruled in the name of some abstract societal value without the presence of concrete harm implications for others.

Conflicting interests and guidance on how to proceed in such conflicts differ according to the type of intervention at stake. We must therefore discuss the main ethical aspects of the various intervention types and consider empirical evidence on outcomes of interventions, in order to satisfy the expectations of a consequential ethics.

Diagnostic Approaches and Procedures

All interventions must be based on a proper diagnosis. This is one of the main basic rules since antiquity. Contemporary medicine has a range of diagnostic tools and techniques that require the informed consent of a patient. Consultation of former medical documents and observations also requires consent. Any diagnostic intervention without informed consent is only admissible if the results are expected to be of relevant practical value for therapeutic planning or for forensic purposes (e.g., in emergencies and in criminal investigations). A diagnosis of addictive behavior has a stigmatizing effect, often with far-reaching consequences. Keeping the medical secret is essential, but often in conflict with the interests of employers, relatives, or others. Specific problems arise if illegal substances are involved or in specific situations (e.g., roadside examination when driving under the influence of addictive substances is suspected).

Targeted Prevention

Universal prevention against substance use and abuse is guided by sociocultural beliefs and increasingly by evidence-based strategies having the intended effects. Targeted prevention is directed toward persons with a high risk to develop addictive behavior (selective prevention) or showing first signs of such behavior (indicated prevention). These types of preventive action have a labeling effect on the targeted persons. Stigma and discrimination even marginalization

are the inherent risks of such actions. If applied without informed consent, they create an ethical problem, even more so in the case of adolescents when parents are involved.

The touchstone is the effectiveness of targeted preventive actions. Weighing the risk of stigmatization against the probability of attaining the intended effects is inevitable. The situation is even more difficult if there is no evidence yet on effectiveness of planned actions.

Targeted action against risk factors, e.g., against adverse childhood experiences, shifts the risks of stigmatization from the child or adolescent to parents and families, and the basic problem with weighing the risks against (proven or unproven) effectiveness is the same. There is no easy solution to the conflict. The easiest cases are individuals or families where stigmatization and marginalization already are present. If not, interventions must be designed in a way that minimizes the negative effects. This effort needs more attention in the future.

Therapeutic Interventions

A hierarchy of objectives. The goals of treatments in substance dependence have changed over the last decades. Traditionally, abstinence was on top of the list. At present, the primary goal is the patients survival, followed by health improvements (or at least avoidance of deterioration), by improvements in social integration, by reductions in substance use (moving away from addictive behavior), and by improvements in quality of life (as defined subjectively by the patient), ultimately resulting in a responsible and satisfactory lifestyle. Abstinence is not conditional for reaching these objectives, nor does abstinence guarantee to reach them.

Tailoring treatment to individual needs. In view of the diversity of etiology, symptoms, and stages in substance dependence, treatment cannot be uniform for all dependent persons. Treatment needs differ between age groups and among other target groups (gender, ethnicity, comorbidity). Treatment must respond to the specific needs of an individual patient on the basis of a comprehensive needs assessment and a treatment planning process where patient and therapist work together on a shared understanding of what is needed and what should be done. Needs-based treatment has better outcomes in comparison to standardized programs, and even a careful assessment of individual needs at entry is gratified by better outcomes. Covering the needs for psychiatric care and living conditions (housing, jobs) is especially important for facilitating a reduction in substance use. “The combination of treatment components and services to be employed must be tailored to meet the needs of the individual, including where he or she is in the recovery process” is therefore one of the principles of addiction treatment (NIDA 2008).

Current Treatment Guidance

Therapeutic approaches and methods cover a large spectrum of pharmacological, psychotherapeutic, and psychosocial interventions. Evaluation efforts have accumulated an increasing body of research evidence on effectiveness and effectivity. Rigorous reviews and meta-analysis of the evidence (e.g., by the Cochrane Collaboration and the Campbell Collaboration), as well as advancements in research methodology (consensus on grades of evidence, on statistical analyses, on qualitative research), are the basis of evidence-based comprehensive guidelines for treatment. In addition, general principles of substance abuse treatment were developed (NIDA 2008). The ethical standard of treatment depends on a good knowledge and rigorous application of this evidence based guidance.

In addition, service requirements include procedural rules and standards of infrastructure. Excluding patients on the basis of their religious or ethnic affiliation is not admissible. Services must assess and respond to all needs of patients, within their organization or by networking with other services. Special attention must be paid to the care of patients suffering from somatic and/or psychiatric comorbidities. Informed consent for all procedures and confidentiality must be standard, as well as nondiscriminatory attitudes and professional competence. Safety and hygiene of premises are essential. Services are accountable to patients, third parties, and the community served.

Substitution Therapies

A special approach is available for selected types of substance dependence. The famous physician Galen prescribed opium to an opium-addicted Roman emperor. Heroin prescribing is British practice since 1920, experimentation with morphine occurred in the USA and other countries, and replacement therapies started in Canada and in the USA in the 1960s. The main objective was to avoid excessive use as well as withdrawal syndromes, by implementing a regime of externally controlled use for those who had lost control over their consumption.

Today, the most prominent and well-researched example is the opioid maintenance therapy (OMT) for heroin-dependent persons, replacing illegal opiates by agonist medications (methadone, morphine, pharmaceutical heroin) or medications with agonist/antagonist properties (buprenorphine, buprenorphine/naloxone). In contrast to the traditional British practice to hand our opiate prescriptions, the prevailing practice today is a supervised intake of medication, in the framework of a comprehensive assessment and therapeutic program. In its review of the evidence, World Health Organization concluded that OMT is the most beneficial treatment approach to heroin dependence and made detailed recommendations on its implementation (WHO 2009). The benefits go beyond the avoidance of the risks in uncontrolled opioid use, providing substantial chances for improvements in health and social integration. However, OMT still meets severe restrictions or even strong opposition in a number of countries.

Another example is the replacement of smoked tobacco by nicotine patches, chewing gums, or non-smokable tobacco. The main objective here is to avoid the cancerogenic properties of tobacco smoke. This type of substitution is rather harm reduction than treatment.

Harm Reduction Approaches

Harm reduction includes all interventions designed to protect or improve the health and social status of chronic addicts who are unable or unwilling to discontinue their addictive behavior. The range of interventions include:

- Needle and syringe availability for drug injectors (to prevent HIV infection through contaminated syringes)
- Safe consumption rooms (to prevent fatal overdose and to provide medical and social care)
- Provision of opiate antagonists to families and peers of opiate injectors (to prevent fatal overdose by emergency medication)
- Drug testing and counseling in nightlife recreational substance use (to prevent harm from unknown substances)

The ethical debate on treatment and harm reduction is often a debate on opposing principles of action. In this debate, harm reduction is disqualified as an approach to prolong dependence, to make substance use acceptable for young people, and to undermine the readiness of addicts for treatment. None of these concerns were substantiated in research, and evaluation resulted in accumulated evidence for good goal attainment of the various approaches (Rhodes and Hedrich 2010). Today, harm reduction is considered an ally rather than an opponent of treatment.

Rehabilitation and Recovery

Substance dependence develops frequently on the background of social and personality factors that make it difficult to conduct a satisfactory life; it also leads frequently to a deterioration of life conditions. Treatment and care cannot be restricted to a reduction of addictive behavior. Patients should be helped to conduct a subjectively satisfactory life in the community. This includes specific rehabilitation efforts, programs, and services, such as supplementary education, vocational training, supported employment, housing, financial, and legal support. A recent movement under the label recovery calls for intensified efforts to rehabilitate addicted patients and enable them to become model citizens.

In an ethical perspective, it is essential to respect the individual potential and the subjective readiness of patients to engage in such a process.

Appropriate assessment of a person's motivation, realistic opportunities, and available options is more acceptable and efficient than imposing a program and a standard that does not respond to the patient's preferences. Refusing the demands of rehabilitation and recovery is a right to be respected, and a new kind of stigmatization for those who prefer to have their own ways (as far as compatible with the rights of others) must be avoided.

Coercive Care

As in psychiatry and in the case of dangerous contagious diseases, there are legitimate boundaries to the liberties of addicted persons, on the basis of rightful respect for the interests of others. Nonvoluntary hospitalization and treatment without informed consent is ethically acceptable, if no other measures are in place to protect a person from self-harm or others from the person's harmful behavior. The rules to be followed in such situations concern a clear responsibility, eventually shared decision making, a careful documentation of reasons, and extent of the measures taken.

Using nonvoluntary confinement and regimes in order to enforce abstinence, as still practiced under medical or law enforcement responsibility in a number of countries, for an estimated 300,000 people, gravely disregards ethical principles and fails to reach desired outcomes (UN 2012). In many other countries, harm reduction measures and addiction treatments (detoxification, therapeutic communities, substitution therapy) are available for convicted addicts. Such interventions must be optional as an alternative to a regular prison regime, in the prison milieu or on court order outside of prison in a community-based service. In-prison services must follow the same rules as community-based services and can be equally effective.

Milder forms of coercion (threats of losing a job, a financial support, the spouse) happen in situations where other measures fail and are typical examples of ethical conflict between interests of patient and others. Acceptability depends on failed attempts to find alternatives or on an agreement with the addicted person. The latter is the basis of a therapeutic intervention called *contingency management* (the patient agrees on defined consequences, e.g., losing driver license, notification of employer if breaking the therapeutic contract). As a rule, positive consequences of reaching a defined goal are more effective than negative consequences of missing a goal.

Requirements at the System/Network Level

Coverage of treatment needs is one of the priorities in a public health perspective, to offer treatment to all persons in need of treatment. Individual care may be optimized by high-quality treatment, but public health cannot accept high-quality standards for a few as long as the many are not reached adequately. Good access to treatment asks for a range of qualified services of different types, easily reached by public transport, open for all in need without discrimination. The range of services must include detoxification, long-term drug-free treatment, opioid substitution treatment, rehabilitation programs, and harm reduction approaches; it also must offer early brief interventions in general medical and social services. Psychosocial assistance, psychopharmacology, and behavioral psychotherapy are also essential elements.

The care system must be an integrated system that enables therapeutic and harm reduction services to work together, in order to provide a continuum of care, including:

- Easily accessible low threshold services that meet the immediate needs of active drug users
- Clear processes for motivating users to move away from drug dependent lifestyles
- Clear processes for referring users into structured treatment programs that promote stabilization or abstinence

This principle includes a monitoring of the treatment needs in a given population and, accordingly, a careful planning of the treatment system as a whole, in response to the identified needs. Ethical and professional responsibilities are identified at multiple levels: medical practitioners are responsible for good individual care, service

directors are responsible for good practice and continued education in their services, and health authorities are responsible for good and cost-effective coverage of treatment, for appropriate regulations and resources.

Research

The World Medical Association Declaration of Helsinki of 1964 (repeatedly amended) formulates ethical principles for medical research involving human subjects. The principles fully apply in research on substance abusers and substance-dependent persons. Special attention is due to the vulnerability of this target group for stigmatization and for legal prosecution when illegal substances are involved. In the interest of good coverage and best use of available resources, the clinically preferred randomized controlled trials on the short-term efficacy of specific methods must be complemented by prospective observational studies, providing evidence on effectiveness and cost-effectiveness (including acceptability, retention, and mid- to long-term outcomes).

Conclusion

The risks for developing substance dependence and other forms of abuse are increasingly well researched. Promising interventions are available, at the individual and at the population level. Ethical rules, typical conflicts, and ways how to deal with those exist in some detail. However, there is much need for a better implementation, in the education of concerned professionals, in health policy, and in service provision. One of the main barriers is the stigmatization of abusers and of those who care for them.

Cross-References

- ▶ Addiction
- ▶ Coercion
- ▶ Human Rights
- ▶ Medicine and Ethics
- ▶ Moral Theories
- ▶ Public Health

References

APA. (2013). *Diagnostic and statistical manual of mental disorders DSM-5*. Arlington: American Psychiatric Association Publishing.

Council of Europe. (1950). *European convention on human rights*. Amended 2010. https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/Convention_ENG.pdf. Accessed 19 May 2015.

Gowing, L., Ali, R. L., Allsop, S., Marsden, J., Turff, E. E., West, R., & Witton, N. (2015). Global statistics on addictive behaviors: 2014 status report. *Addiction, 110*, 904–919.

Murray, C. J. L., & Lopez, A. D. (1997). Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet, 349*, 1436–1442.

NIDA. (2008). *Principles of addiction treatment. A research-based guide* (3rd ed.). Rockville: National Institute of Drug Abuse. revised 2012.

Rhodes, T., & Hedrich, D. (Eds.). (2010). *Harm reduction: Evidence, impacts and challenges*. Lisbon: European Monitoring Centre for Drugs and Drug Addiction.

UN. (1948). *The universal declaration of human rights*. http://www.un.org/en/documents/udhr/. Accessed 19 May 2015.

UN. (2012). *Compulsory drug detention and rehabilitation centers*. A joint statement of 12 UN agencies.

UNODC. (2015). *World Drug Report 2014*. http://www.unodc.org/documents/data-and-analysis/WDR2014/World-Drug_Report_2014_web.pdf. Accessed 19 May 2015.

Van Amsterdam, J., Nutt, D., Phillips, L., van den Brink, W. (2015). European rating of drug harms. *Journal of Psychopharmacology* 1–6. doi:10.1177/0269881115581980.

Volkow, N. D., Fowler, J. S., & Wang, G. J. (2003). The addicted human brain: insights from imaging studies. *Journal of Clinical Investigation, 111*, 1444–1451.

West, R. (2006). *Theory of addiction*. Oxford: Blackwell Publishing.

WHO. (2002). *World health report*. Geneva: World Health Organization.

WHO. (2009). *Guidelines for the psychosocially assisted pharmacological treatment of opioid dependence*. Geneva: World Health Organization.

WHO ICD-10. (1990). *The ICD-10 classification of mental and behavioral disorders. Clinical descriptions and diagnostic guidelines*. Version 2015. World Health Organization, Geneva.

Further Readings

Helmchen, H., & Sartorius, N. (Eds.). (2010). *Ethics in psychiatry*. Dordrecht: Springer Publishers.

UNODC/WHO. (2008). *Principles of drug dependence treatment.* Discussion paper. http://www.unodc.org/documents/drug-treatment/UNODC-WHO-Principles-

of-Drug-Dependence-Treatment-March08.pdf. Accessed 19 May 2015.
WHO. (2010). *ATLAS on substance use – Resources for the prevention and treatment of substance use disorders*. Geneva: World Health Organization.

Suffering

Jeff Malpas
University of Tasmania, Hobart, TAS, Australia
Latrobe University, Bundoora, VIC, Australia

Abstract

The topic of suffering encompasses an enormous range of issues. In this discussion the focus will be is on two key sets of questions: the nature and definition of suffering and the nature of the responses to suffering. At its broadest, suffering is taken to be identical with any negative or "unpleasant" experience, but such a conception suffering gives rise to several problems. On a more nuanced account, suffering is tied to a disruption or potential disruption to the integrity of the person. The differences between these two accounts are themselves tied to differences in responses to suffering and in approaches to the relief of suffering. As it is central to any attempt to understand human being in the world, the topic of suffering is central within bioethics but also a challenge to it.

Keywords

Bioethics; Compassion; Distress; Meaning; Narrative; Pain; Person; Self/other; Suffering; Relationality

Introduction

The topic of suffering is one of the most profound and far reaching. It brings to the fore a set of key issues concerning the character of human being, of the relations between human beings, and of the relation of human beings to the world (including the relation to other nonhuman beings). Two sets of questions come to the fore in most discussions of suffering within bioethics: first, questions concerning the *nature and definition of suffering*, including the experience and significance of suffering, and, second, questions concerning the nature of the *response to suffering*, including the imperative to relieve suffering and the implications of that imperative and the manner in which it is realized. These questions inevitably reach back to the fundamental issues concerning human being that are at stake here, while they also extend outward to encompass a range of further issues including the physiology and psychology that underpins suffering (often addressed in connection with the physiology and psychology of pain); inequalities in suffering across societal, cultural, and geographical divides; the appearance of new forms of suffering that may be consequent on social, technological, or political change; the interrelation between suffering and discrimination or between suffering, justice, and law; the cultural and historical context in which suffering is understood; and the portrayal of suffering in its personal, institutional, and historical dimensions. The extent of these questions is enormous, and the discussion in this chapter cannot do justice to all of them. Instead the aim here is to provide an analysis of the nature of suffering and of the response to suffering in terms of the core philosophical and conceptual issues at stake.

History and Background

The history of human attempts to express and articulate responses to the experience of suffering goes back to the very earliest forms of recorded human thought. Whether one looks to the Babylonian *Epic of Gilgamesh*, the Jewish *Torah*, the *Iliad* and *Odyssey* of Homer, the *Analects* of Confucius, or the Christian *Gospels*, suffering appears as a central and persistent theme. There is thus a history of thinking about suffering, including the relation between suffering and ethics, that long precedes contemporary bioethics (see Amato (1990) for an account of suffering in its broader historical and political context). Yet

notwithstanding the history and importance of the topic, suffering has not always received as much attention within bioethics as have many others. This is undoubtedly a consequence of the very breadth and depth of the topic itself but also of the tendency for bioethics to focus on more specific and immediately contentious topics often directly related to specific problems of technology or practice. Moreover, traditional bioethical thinking has also been heavily influenced by models of analysis derived from a natural scientific model (not surprisingly given the dominance of such models within contemporary medicine) that have often seemed ill equipped to deal with the complex character of suffering and especially its embeddedness in the experiential and the "subjective." The last 30 years or so have nevertheless seen increasing attention being given to the topic of suffering in the bioethical and biomedical literature. This has been due to a number of factors, including the rise of palliative medicine, of more patient-oriented approaches to pain, and the greater influence, within bioethics, of philosophical approaches from within the phenomenological, hermeneutic, psychoanalytic, and poststructuralist traditions.

The Nature of Suffering

Suffering, Pain, and Distress

Etymologically, suffering comes from the Latin *suffere*, meaning to bear or undergo. In this respect, it is clearly related in meaning to the Greek *pathe*, meaning to be affected by, which is itself related to other Greek terms meaning suffering, pain, and grief, and also to the Latin *passio*, which itself means suffering and from which comes the English *passion*. In its simplest and most everyday usage, suffering is experiencing or being affected that is negative in character. Understood in this broad sense, suffering is indistinguishable from pain, whether occasioned by physical or mental hurt, and may be borne individually or collectively; in this sense, it encompasses all forms of hardship and distress.

It is something like this broad conception of suffering (though understood primarily in terms of the suffering of individuals) that Peter Singer argues is the basis for moral concern. To be a creature that has interests worthy of moral concern, claims Singer, is to be a creature that can experience suffering or, on the positive side, enjoyment (Singer 2011: 50). Given his utilitarian standpoint, Singer's conception of suffering is one that allows suffering to be quantified: suffering is increased as the number of individuals who suffer increases, and, in some case, suffering increases according to the capacities of the individuals who suffer (whether because of greater physical vulnerability, sensory sensitivity, or capacity for emotional and cognitive response). Although Singer is an influential figure in contemporary bioethics, his account of suffering has not figured significantly in the biomedical accounts of suffering. It nevertheless deserves notice not only because of Singer's wider influence, but also because that account does indeed seem to fit with the idea of suffering as broadly understood, and because it understands suffering as something *felt* rather than merely *inflicted*. Singer's account of suffering, which is essentially an account of suffering as "negative experience," echoes in the background of some other accounts, notably Jamie Mayerfeld's definition of suffering as "disagreeable overall feeling" (Mayerfeld 2005: 14). Mayerfeld characterizes "suffering" as *psychological distress* and so, in his terms, as "subjective" rather than "objective."

That suffering is indeed "subjective in the sense that it pertains *to the subject's experience of* the world (rather than pertaining to the world as it might be construed *apart from* the subject's experience of it) seems clear. That an individual suffers is thus a fact *about their experience*, rather than being determined by anything independent of that experience. The same can be said to be true of pain, and yet the tendency of broad accounts of suffering, including Singer's, to treat suffering as more or less indistinguishable from pain gives rise to problems. One reason for this is that there are surely instances of pain that do not count as suffering. One might say this is true of momentary pain, and perhaps of most pain of relatively short duration, or of pain that is not severe but also of pain to which we stand in a particular cognitive or

emotional relation. The pain one may feel in accomplishing something – say the pain of an athlete as she pushes her body in competition – is not, by the mere fact that it is pain, also to be counted as suffering. Equally, the pain one may feel when undergoing some medical procedure, assuming it is properly administered with a therapeutic intent, is not always taken to be an instance of suffering. There are clearly cases of what might colloquially be referred to as mental pain or discomfort – whether incurred in the process of completing some bureaucratic process or trying to think through a difficult conceptual problem – that are not, in virtue of their being extremely discomforting or even, in some sense, "painful," also cases of suffering.

On Singer's account, pain can certainly be outweighed, in a utilitarian calculation, by an associated good – and these latter cases could be viewed in this way: the therapeutic goal being primary in the case of the medical procedure or the achievement, or hope of achievement, in the athletic case (and analogously so in the mental case). Yet precisely because Singer's account is a quantitative and cumulative one, so it seems that the pain still remains as itself a form of suffering. Utilitarianism aside, one might acknowledge that to experience pain may well be to suffer, at least in some cases, but also claim that suffering is a matter of the overall experience of which pain may be but a part, as in Mayerfeld's talk of "overall experience," and not of any separate part of the experience (if we can indeed attach sense to this). Equally one might distinguish between different kinds, degrees, or levels of pain – only some of which, perhaps the most severe, count as suffering – and between different kinds, degrees, or levels of suffering also. Ronald Anderson argues that suffering subsumes pain or, at least, severe pain but also distinguishes between three different kinds of suffering, physical, mental, and social, with pain falling into the first (Anderson 2014: 2–3).

Suffering and the Person

In allowing that there can be forms of suffering that go beyond pain alone, Anderson follows a tendency that has largely prevailed across the bioethical literature. This suffering has generally been seen as a form of severe and encompassing distress that is not simply reducible to pain and that may even be said to affect the very character of a life in its entirety. Such an encompassing conception is evident in the now-classic definition of suffering advanced by Eric Cassell, according to which suffering is a state of severe distress in which the integrity or intactness of the person is threatened (see Cassell 2004: Chap. 3). Although it implies a distinction between suffering and pain, as well as between suffering and distress (even though suffering may be said to be *a form* of distress, it cannot be said to be *identical with* distress), Cassell's emphasis on suffering as pertaining to the person, and so to the "whole" (since the person is a whole), as well as his treatment of suffering as "subjective" (a state that is felt or experienced by the sufferer) can be seen as convergent with Mayerfeld's emphasis on suffering as an "overall feeling." Where Cassell's definition differs most significantly from Mayerfeld's, however, is precisely in the way it draws upon the notion of personhood and so implies a degree of self-referentiality.

As Cassell emphasizes, what is threatened in suffering is indeed the intactness of the person *as a person*. Suffering is thus, on Cassell's account, what one might call a *self-reflexive* or *self-regarding* concept – it is to have a sense of *one's own* impending disintegration. Mayerfeld's definition, like Singer's use of the notion, requires no such sense of personhood or of self. One might argue that, in this respect, Cassell's definition of suffering is too narrow since it appears to rule out suffering where there is no capacity for self-regard or self-reflection – where there is no sense of impending personal self-disintegration (as in the case of nonhuman animals or human infants). Rather than Cassell's definition being too narrow, however, it should rather be understood as itself encapsulating the importance of the capacity for self-regard or self-reflection in making possible a distinction between pain or distress and suffering. It is not the case that those creatures that lack an ability to articulate a sense of self that is threatened in the experience of suffering do not suffer or that they suffer less, but rather that in such cases,

the pain and distress at issue are identical with the suffering. Consequently, there is no need to choose between a narrow and broad sense of suffering – accepting one and rejecting the other. Instead, both senses can and should be acknowledged. Significantly, the lack of a sense of self or personhood on the part of the creature that suffers does not itself imply any lessening in the ethical importance of the suffering at issue nor, taken on its own, does it justify any lessening in the consideration that it demands – indeed, one might argue that in some cases, it indicates the need for a greater degree of attentiveness and responsibility.

The Complexity of Suffering

The way suffering, understood as distinct from pain and distress, is tied to the idea of impending personal disintegration is indicative of the complex nature of suffering (and the complexity of the pain and distress that may be associated with it), once suffering is indeed distinguished from pain or distress understood more broadly. Just as there is enormous variation between persons and between different persons' experience of the world – differences that are underpinned by differences in cultural and social background, education and occupation, family, and character, to name but a few – so there is enormous variation between the suffering that persons undergo and between the personal circumstances that may give rise to suffering. The complexity of suffering thus reflects the complexity of "personal," that is to say, of *human* life (where "human" is used to indicate not a biological category but an *ontological* or *ethical* one). The recognition of such complexity is evident at many places in the existing bioethical and biomedical literature.

The implication of the person and the self in suffering means that in the experience of suffering, one is brought face-to-face with one's own being as a person. Existential and phenomenological analyses often tend to emphasize the importance of *death*, or the anticipation of death, as that which brings with it a sense of the personal character of existence – that is, it makes salient one's being as one's own – and so makes salient the way in which to exist is precisely for one's being to be an issue for one. Yet it is perhaps the real experience of suffering that is actually the key here, since that suffering, understood as an experience of impending personal disintegration (which it may be argued is more profound than the mere intellectual anticipation of one's own demise), is indeed that which brings our own personal being directly and unavoidably into view – that does indeed most directly and concretely make it into an issue for us. Like the anticipation of death, but again perhaps, more concretely, suffering also makes salient the essentially finite character of existence (and this may be the key to the character to existence as *personal*) – that is, its essential formation in relation to *limit* – and, on this basis, suffering must be understood as a necessary, and not merely contingent, feature of life and existence. The latter point is contested, however, by an abolitionist strain in contemporary thought, partly deriving from utilitarian thinking, and closely associated with transhumanism, which argues for the elimination of suffering by technological means as a necessary and achievable end. The abolitionist position aside, however, there is also a tradition of thought that takes the connection between suffering and limit to be indicative of the importance of suffering to the possibility of knowledge (where knowledge is itself knowledge of both limit and of self) and so as closely connected with the injunction "suffer and learn" (*pathei mathos*) that appears in Aeschylus' *Agamemnon* (see Aeschylus 2009: line 177).

The personal character of suffering may suggest that suffering is always an experience that is deeply introverted – that it is an experience that turns back toward the self in which the sufferer loses sight of any connection to others or to the wider world. Yet although there is a sense in which suffering can be construed in this fashion (partly because suffering can itself undermine the usual sense of the distinction of self from world so all that exists is the "subjective" world that is one's suffering), it cannot be taken as an adequate characterization of what occurs in suffering or in the experience of suffering. Central to Cassell's definition of suffering, and one of its central insights, is a conception of the person as complex and essentially *relational*. The threat to the person

that occurs in the experience of suffering is a breakdown in the relationality of the person – in the capacity to integrate the elements of the person. Although it always involves *self*-relationality (since the sense of impending breakdown itself involves a mode of self-reflexivity), such breakdown can arise with respect to any of the relations by which personhood is articulated and often encompasses many or all of them. Moreover, the self-relationality that is at issue here is not a relationality to a self that remains itself secure – suffering is precisely a state in which the self is under threat, in which we lose a sense of orientation to ourselves, to others, and to the world. It is thus that Elaine Scarry can indeed speak of the suffering that is felt in certain experiences of pain and especially of the extreme pain and cruelty evident in torture, as an "unmaking of the world" (see Scarry 1986).

If suffering forces us to a recognition of our own personal being, then it also has the potential to bring to light our own being with others – even if this is evident in its incipient breakdown. Certainly there is the potential for the sufferer to respond to suffering in anger or bitterness, withdrawing into their suffering in a way that severs them from themselves as well as from others, but suffering also carries within it the potential to open us up to others through the very relationality that it brings into focus – as Tolstoy might be taken to suggest in the conclusion to *The death of Ivan Ilyich* in which Ivan is opened to those around him and so released from his imprisonment in his own anger and hatred, in a way that also provides release from the suffering that he has endured (Tolstoy 1989 [1886]).

The capacity for suffering to open us to others, and to our relationship to others, is relevant, moreover, not only to those who suffer but also for those who are witness to suffering – at least if they allow themselves to be drawn into that suffering. What suffering has the potential to lay bare, then, is the character of human or "personal" being as not merely a being with oneself but always a being with others as well as within a world. It is not that being oneself comes first, and that from there one moves out to others, but rather that being oneself is only possible in relation to being with others. Such a notion, though variously articulated, can be found in philosophical form in the work of such twentieth-century thinkers as Max Scheler, Martin Buber, Hans-Georg Gadamer, and Emmanuel Levinas. In Levinas' work, in particular, the idea of the interpersonal or "the interhuman" (to use Levinas' term) is developed in explicit relation to the ethical demand placed upon us by the encounter with others and especially by the encounter with the other as one who suffers (Levinas 1998 [1982]).

Accounts that tie suffering to the idea of the person stand in sharp contrast to those, like Singer's, that treat suffering as identical with pain or distress – as "negative experience" – and this contrast is not merely one of breadth or narrowness in definition. Understood as the experience of a threat to the very being of the person, there is an important sense in which suffering stands outside of any "calculation" and may even be said to resist the very possibility of rational interrogation. Suffering thus appears as a "singularity," such that it may even be said that suffering reaches its absolute in any and every instance of an individual who suffers. Certainly, on such an account, suffering is not divisible, and the pain and distress that may be discerned in suffering are not separate instances of suffering, since suffering is an experience *of the person*, rather than of any "part" of the person. If, on the other hand, one takes suffering to be identical with pain or distress in general, then one is committed to the view that any discrete instance of pain or distress is also an instance of suffering, as well as to the possibility that suffering can be understood as composed out of other instances of pain and distress – out of other instances of suffering. Suffering is then both potentially divisible and cumulative, and it is also potentially amenable to quantification and calculation, whether for an individual who suffers or across many such individuals.

The latter conclusion is an obvious component in Singer's approach – it is a fundamental element in his commitment to utilitarianism. Yet whether or not it is formulated in terms of an explicit utilitarian commitment, it is also a common element in much contemporary bioethical thinking for the simple reason that it does indeed enable

suffering to be accommodated within a rational scheme – within a possible "economy" of suffering in which suffering can become the subject of assessment and calculation. The way in which it takes suffering to stand outside of any such "economy" is part of the genuine radicality of Cassell's account of suffering, a radicality only partially realized in his own development of that account and often passing unremarked upon in the way that account is taken up. Genuinely taking account of the personal and singular character of suffering almost certainly implies a rethinking of many aspects of contemporary medical practice (including the intrusion into medicine of neoliberal models of management and decision-making) and even of current bioethical orthodoxy. It also makes the task of responding to suffering as complex as is the phenomenon of suffering itself.

Responses to Suffering

The relief of suffering has been at the heart of medical practice from the beginning, but one might also argue that it lies at the core of ethical conduct when understood in explicitly relational terms. One might even argue that suffering carries within it an ethical demand that is absent in the experience of happiness: there is an ethical demand to relieve suffering, but no analogous demand to increase happiness. Utilitarians would dispute the seeming asymmetry between suffering and happiness that appears to emerge here, but even if it is rejected, and suffering and happiness are seen as standing exactly counter to one another and of equal ethical weight, still the relief of suffering remains a powerful ethical imperative. For utilitarians, it is simply that suffering is not exclusive in this regard (although the "emotional" power of suffering could also be seen as potentially misleading and as tending precisely toward an overestimation of the ethical significance of suffering in relation to happiness).

Suffering remains an important moral concern regardless of whether one takes a utilitarian or nonutilitarian approach. Nonetheless, exactly how one understands suffering may have important consequences for what one takes to be the best means of responding to suffering or to what one takes as the best means of achieving the relief of suffering. If suffering is understood, for instance, as identical with pain, then the relief of pain must constitute the relief of suffering, and if pain could be universally eliminated, then so too would suffering be universally eliminated along with it – a utopian ambition embraced by some, as noted above, but warned against by others. Yet if suffering is understood as distinct from pain, and instead as tied to the threatened disintegration of the person as in Cassell's account, then the relief of pain alone may be insufficient for the relief of suffering, and the capacity to relieve pain may also turn out to be limited in a way that parallels the character of suffering as itself a marker of the limitation or finitude that is itself a central feature of the existence of persons. Understood as a condition pertaining to the person, suffering can only be addressed by addressing the person and the overall conditions affecting the person. Moreover, not only does this require more than addressing pain alone, but because it requires attending to a holistic complex of elements, so it will never be amenable to any complete or determinate control. On this account, there can be no "technology" for the elimination of suffering, since there can be no "technology" of persons.

The emphasis here on the need, if one is properly to address suffering, to address the person "as a whole" and so to attend to the complexity of personal life and situation is itself captured in the emphasis on the character of suffering as involving a breakdown in the possibility of "meaning" or "significance" (something already suggested by the idea of suffering as involving a "loss of orientation" of the self). The meaning or significance at issue here is precisely the meaning or significance associated with a sense of personal integration or identity, and this sort of meaning or significance typically takes the form of a capacity to find *narrative* structure in one's experiences and in the events in which one is involved. There is considerable philosophical and psychological literature that takes the self to be constituted in such narrative terms, and similarly the bioethical literature contains many discussions of the importance of narrative in the response to

suffering (see, e.g., Carr et al. 2005). Cassell notes that "assigning meaning to [an] injurious condition often reduces or even resolves the associated with it" (Cassell 2004: 43), and this is precisely what narrative enables. Cassell also notes that as it involves a sense of "impending" disintegration, so suffering has an essentially temporal dimension (see Cassell 2004: 35), and this directly implicates ideas of narrative as a projection of future possibility.

The way in which suffering seems to stand opposed to meaning is given a particularly powerful form in the seeming opposition, within many traditional religious contexts, between suffering and the idea of beneficent divinity (whether understood in terms of a divine person or a divinity that attaches to the universe as a whole). Indeed, this opposition can almost take the form of a dilemma: on the one hand, the experience of suffering can impel us toward a notion of the divine as a means to uncover meaning (indeed, suffering plays a key role in religious thinking within bioethics as well as more broadly – see, e.g., Engelhardt 2000), while on the other hand, the very existence of suffering seems to be incompatible with the existence of the divine (or at least with the idea of divinity as beneficent).

In Albert Camus' novel *The Plague*, the priest Paneloux and the doctor Rieux both witness the suffering and death of a child. The priest's response is that what they have witnessed is "beyond us," but then he goes on to add that "perhaps we should love what we cannot understand." The response of the doctor, Rieux, which undoubtedly coincides with Camus' own, is quite contrary: "'No, Father,' he said. 'I have a different notion of love; and to the day I die I shall refuse to love this creation in which children are tortured.'" (Camus 1971 [1947]: 178). The episode not only echoes an idea found elsewhere (most notably in Dostoyevsky 2002 [1880]: Chap. 5), but it also exemplifies Camus' opposition of medicine, as that which fights against suffering, to the divine, as that which sanctions suffering (an opposition that derives less from Camus' atheism as his profound "humanism"). Here what also comes into view (and is especially clear in Dostoevsky) is the question – a pressing one within medical practice itself and whose answer is once again dependent on how suffering is understood – as to whether and under what circumstances it can ever be acceptable to inflict suffering on another (Camus rejects, as does Dostoevsky, the idea of justifying the suffering of one by the welfare of others while remaining silent on the matter of suffering inflicted with the consent of the one who suffers). Camus' position represents an extreme version of the traditional problem that suffering, as the paradigm case of evil, presents for theology, and to which theodicy is a response. Yet it also stands as another example of the centrality of suffering to any attempt to think about the world and human existence – suffering not only presents a problem for the idea of a beneficent divinity (to which the idea of the divine may also appear as an answer) but for the very attempt to make sense of human being in a world so given to contingency and seeming senselessness.

Responding to suffering, whether in specific instances or more generally, requires that we find ways to reconstitute the meaning that suffering seems to rend asunder, and it is here that narrative plays a central role. Yet narrative is also significant as one of the means by which sufferers are able to give expression to their suffering. The very character of suffering as threatening both personal integration and the possibility of meaningfulness (which here amount to the same thing) is indicative of the way suffering can also threaten the very capacity for expression and communication, and the consequent isolation itself becomes an additional source of suffering. In Sophocles' play, the Greek hero Philoctetes, stricken with a festering wound that leads his fellow Greeks to abandon him on a deserted island, is unable to speak his suffering other than through inarticulate cries (Sophocles 1994). In the worst throes of his suffering, Philoctetes' exile from others is also a seeming exile from language. From a medical perspective, finding ways to enable a patient to express and communicate their suffering is surely a prerequisite for successful medical diagnosis and treatment, but more than this, it may be an essential component in the treatment itself.

The imperative to relieve suffering is frequently cited as a key consideration in the

argument for euthanasia – or what is often termed "physician-assisted suicide" or (somewhat euphemistically) "physician-assisted dying." Yet here too, much depends on how suffering is itself understood. It is sometimes claimed that the imperative to relieve suffering is incompatible with the medical imperative to do no harm or more specifically with the ethical prohibition against killing. But this assumes that the death of the one who suffers is indeed a *relief* of suffering rather than simply a *cessation* of suffering itself consequent on the cessation of the life of the one who suffers. If relief of suffering is more than the mere cessation of suffering (just as it is more than pain or distress), then the imperative to relieve suffering need never come into conflict with the imperative not to harm or not to kill. Moreover, the prohibition against killing might itself be particularly important in medicine given the enormous power of the medical practitioner as against the vulnerability of the patient and especially of the suffering patient (inasmuch as suffering threatens the integrity of the person, then so it also threatens the capacity of the person to maintain their own integrity, and that may well include the capacity for effective decision-making even in relation to their suffering). Where euthanasia is advanced as a remedy for suffering, then it is most persuasively advanced as a remedy in cases where patients are also at their most vulnerable and so in cases where the risks of increased harm or suffering are at their greatest – cases in which the exacerbation of patient vulnerability through suffering becomes itself a reason that counts against euthanasia and for a more cautious and careful approach.

Although the relief of suffering is often cited as a primary consideration in arguments for euthanasia, considerations concerning the nature of suffering and its relief, and the vulnerability of suffering patients, suggest that the argument is more plausibly (but, in some respects, perhaps less persuasively) founded in the assertion of individual autonomy in a way essentially independent of any question of suffering as such. Euthanasia is thus founded in the absolute right of the individual to dispose of their lives however they see fit, but, in addition, it must also make a claim, whether explicitly or not, to a right to be supported in the exercise of that autonomy by others. There is some irony in this, not to say a fairly obvious problem, since it allows the assertion of autonomy to be used in a way that makes claims on others (whether on individuals or the state) in a way that potentially infringes on the autonomy of those others and is often effectively an attempt at the coercion of others to act in ways consistent with the wishes of the individual who makes the claim. Yet regardless of the internal consistency of the arguments used to advance euthanasia on the basis of autonomy, such strong assertions of autonomy seem to be incompatible with the sort of relational view of persons that underpins accounts of suffering such as Cassell's. If persons are understood relationally (in terms both of self-relationality and relationality to others), then the autonomy that pertains to persons cannot be construed as absolute but must always be interpreted against the background of the relationality out of which it comes and so as constrained by that relationality.

One of the ways in which the relationality at issue here is evident in a particularly relevant and significant fashion is in regard to the effect of patient suffering on the experience of the doctors and nurses who care for them. What, one may ask, is the appropriate comportment of medical practitioners to the suffering of their patients – should it be one of "objective" detachment or of compassionate solidarity? There is good reason for arguing that if one accepts that suffering involves the entire person and that the person is indeed a relational entity, then the second of these responses is the more appropriate (and may in one sense be seen as more rather than less objective precisely in allowing greater access to the reality of the patient's experience). As the term itself suggests, however, *compassion* is itself a form of "suffering with," and although this may be important in allowing medical practitioners better to understand their patients, and so to care more effectively for them, it surely also opens practitioners more directly to the patient's own suffering, placing a burden on the practitioner that may not be so obviously present (although it may be argued that it is merely suppressed) in cases where a more "objective and detached" stance is adopted.

Conclusion

The nature of suffering and the responses to it constitute questions central to bioethics, as they are to ethics more generally (and to any genuine attempt to think the nature of the human or of human being the world). Yet how one responds to those questions also makes for radical differences in how bioethics is itself understood and in the manner of the thinking in which it is taken to consist. Moreover, what one takes as the evidence on which those questions are addressed and the sources of that evidence are also key issues – part of the implicit argument here is that not only clinical practice or philosophical analysis but also literary and artistic expression and exploration may be vital in any adequate thinking of suffering and its implications. The questions at issue in suffering not only constitute some of the most basic questions that bioethics must address, but they also present a challenge to bioethics as such (especially to bioethics in more technically oriented forms), since they bring into view what ought to be foundational to both bioethical thought and biomedical practice, as well as their very limits.

Cross-References

- Assisted Suicide
- Autonomy
- Care Ethics
- Death: Good Death
- Death: Meaning of
- Empathy
- Ethics
- Euthanasia: Active
- Euthanasia: International Debate
- Euthanasia: Passive
- Hospice
- Humanitarian Action
- Life: Meaning of
- Literature
- Medicine and Ethics
- Narrative Ethics
- Pain
- Palliative Care
- Person
- Personalism
- Professional Ethics
- Quality of Life
- Religion and Global Bioethics
- Right to Die
- Spirituality
- Torture
- Transhumanism
- Triage
- Utilitarianism
- Utopianism
- Vulnerability

References

Aeschylus. (2009). *Oresteia: Agamenon, Libation-Bearers, Eumenides* (trans: Sommerstein, A. H.). Harvard: Harvard University Press (Loeb Classical Library).

Amato, J. A. (1990). *Victims and values: A history and a theory of suffering*. New York: Praeger.

Anderson, R. E. (2014). *Human suffering and quality of life: Conceptualizing stories and statistics* (SpringerBriefs in well-being and quality of life research). Dordrecht: Springer.

Camus, A. (1971 [1947]). *The plague* (trans: Gilbert, S.). London: Penguin.

Carr, D. B., Loeser, J. D., & Morris, D. B. (Eds.). (2005). *Narrative, pain, and suffering*. Seattle: IASP Press.

Cassell, E. J. (2004). *The nature of suffering and the goals of medicine* (2nd ed.). Oxford: Oxford University Press.

Dostoyevsky, F. (2002 [1880]). *The brothers Karamazov: A novel in four parts with epilogue* (trans: Pevear, R., Volokhonsky, L.). New York: Farrar, Straus and Giroux.

Engelhardt, H. T., Jr. (2000). *The foundations of christian bioethics*. Lisse: Swets and Zeitlinger.

Levinas, E. (1998 [1982]). Useless suffering. In *Entre Nous: On thinking-of-the-other* (pp. 91–101, trans: Smith, M.B., Harshav, B.). New York: Columbia University Press.

Mayerfeld, J. (2005). *Suffering and moral responsibility*. New York: Oxford University Press.

Scarry, E. (1986). *The body in pain: The making and unmaking of the world*. New York: Oxford University Press.

Singer, P. (2011). *Practical ethics*. New York: Cambridge University Press.

Sophocles. (1994). *Antigone. The women of trachis. Philoctetes. Oedipus at Colonus* (trans: Jones, H. L.). Loeb Classical Library. Harvard: Harvard University Press.

Tolstoy, L. (1989 [1886]). *The death of Ivan Ilyich and other stories* (trans: Edmonds, R.). London: Penguin.

Further Readings

Anderson, R. E. (2015). *World suffering and quality of life* (Social Indicators Research Series, Vol. 56). Dordrecht: Springer.

Bourke, J. (2014). *The story of pain: From prayer to painkillers*. Oxford: Oxford University Press.

Green, R. M., & Palpant, N. J. (Eds.). (2014). *Suffering and bioethics*. Oxford: Oxford University Press.

Malpas, J., & Lickiss, N. (Eds.). (2012). *Perspectives on human suffering*. Dordrecht: Springer.

Suicide

Jann E. Schlimme
Charité Universitätsmedizin Berlin, Berlin, Germany

Abstract

Suicide is a ubiquitous human phenomenon. It can be found in all living human cultures and is already a topic in early works of literature of mankind. Intentional self-killing requires self-awareness and awareness of one's mortality. As far as we know, only human beings display intentional self-killing. A major moral concern regarding suicide in all human cultures addresses the question of whether a suicide is justified or not. From a philosophical point of view, suicide is primarily an indifferent act. Its ethical evaluation draws on the two moral principles of autonomy and welfare and depends both upon personal and situational (social, cultural, or religious) circumstances connected with this very suicide of this person in this special situation as well as on the judging of person's values.

Keywords

Autonomy; Authenticity; Coercion; Evaluation; Indifferent act; Justification; Moral dilemma; Paternalism; Self-awareness; Welfare

Introduction

After discussing historical and religious arguments regarding the morality of suicide, some conceptual clarifications will focus on a distinction of suicide as intentional self-killing and parasuicide as intentionally using the gesture of suicide. Self-awareness will be addressed as necessary condition for suicidal behavior. A more thorough definition of suicide shows liberty as a focus of philosophical debates on suicide and points out questions concerning the moral justifiability of suicide, suicide prevention, and assisted suicide as their major concerns. In the main body of the entry, the ethical dimensions of suicide and the arguments deployed in this debate will be discussed in four subsequent sections: "The Suicidal State of Mind and the Medical Argument of Mental Illness," "Actual Deontological Arguments Against Suicide," "Social Arguments Regarding the Morality of Suicide," and "Can Suicide Be Authentic?" Some short conclusions will be given at the end of each of these sections. The closing section of the entry will offer rather general conclusions concerning the three major philosophical questions regarding the moral justifiability of suicide, suicide prevention, and assisted suicide.

Historical and Religious Arguments Regarding the Morality of Suicide

Understanding and judging suicide has a long, explicit tradition in high cultures since around 2000 BC but may have been performed and debated already earlier. The general connotation of suicide is typically ambivalent. While it is agreed that suicide is a behavioral option, usually taken into account only in a personal crisis, the act itself can either be good or bad. This implies that most cultures and worldviews (including religious faiths) define certain conditions under which suicide is accepted or even glorified. The first text dealing explicitly with this topic is *The Report about the Dispute of a Man with his Ba* (Papyrus Berlin 3024 from 1800 BC). According to the cosmic order during those times (Egypt in the Middle Kingdom), human beings need time in order to prepare for their life after death, while the *Ba* (soul) achieves greater freedom from the human being after death. Hence, suicide in this

life denies the cosmic order if one's death is not prepared adequately. Accordingly the man rejects the *Ba's* offer of an early death.

Since 500 BC more explicit debates on the topic of suicide are handed down to us in different cultures. For example, in the Jewish faith suicide is violating the gift of life, given and taken by *YWH*. Hence, suicide violates the cosmic order of life as God's gift. There are, however, nine cases of suicide in the texts of the *Old Testament* (written between 500 and 200 BC). As is prominently displayed in the case of Samson, suicide seems acceptable if the person's wish to put herself to death is endorsed by God. In this vein, the person's suicide becomes a glorification of God. This raises, of course, the question how God's endorsement is given and correctly perceived by the involved human beings. Since this is most difficult to decide, and cannot simply depend on the suicidal person's perspective, it is understandable that suicide was – and is – usually not accepted from a Jewish perspective. This position is furthermore undergirded by social arguments, especially the argument that one's intentional death does no harm to one's community, society, or state (see below). There are exceptions of an acceptable suicide, as has constantly been argued from more Orthodox Jewish perspectives, drawing, i.e., on the abovementioned cases in the *Old Testament*. This complex moral position – suicide is morally unjustified, except a divine sign has been given – was adopted by Christianity (especially by Paul, being originally a Pharisee Jew himself).

In ancient Chinese philosophy (*Confucianism*) suicide can also be the "right" thing to do under certain conditions. According to the cosmos' (heaven = *tian*) order, the highest value for the "right way of living" is righteousness, ideally acknowledged and adopted by human beings. In order to keep on the right way in one's life, one might be obliged to choose life-endangering behavior or even suicide, as Mengzi (cir. 370–290 BC) argued:

> So, I like life, and I also like righteousness. If I cannot keep the two together, I will let life go, and choose righteousness. [. . .] There are cases when men by a certain course might preserve life, and they do not employ it; when by certain things they might avoid danger, and they will not do them. Therefore, men have that which they like more than life, and that which they dislike more than death. They are not men of distinguished talents and virtue only who have this mental nature. All men have it; what belongs to such men is simply that they do not lose it [righteousness J.S.]. (Mengzi Gaozi I, 10)

In the classical Brahmanian perspective, suicide is usually taken as violating the cosmic order. According to the later texts of the *Dharma-sutras*, such as *Manusmriti* written by the mythical figure Manu between 200 BC and AD 200, suicide is only acceptable if performed in a certain manner (performed through fasting, called *prayopavesa)* and under certain conditions (i.e., elderly person). In all other cases, it is an unseemly act, violates the "right way of living" (*dharma*), and creates bad karma to face in the future. According to the wheel of *samsara*, reincarnation of one's immortal *atma* (soul) takes place in accord with one's karma. Therefore, suicide brings a series of immediate lesser births and requires several lives for the soul to return to the exact evolutionary point that existed at the moment of suicide. In today's Hinduism, this manner of killing oneself is still accepted, while in Buddhism suicide is usually seen as contradicting the path to enlightenment.

If we consider these metaphysical perspectives on the moral justifiability of suicide, three arguments can be distinguished: (a) the property argument, (b) the providence argument, and (c) the cosmic or natural law argument. All three arguments argue suicide as usually not acceptable but specify also acceptable manners of suicide under certain conditions:

(a) The *property argument* claims that our life does not belong to us, but to a different and more powerful entity (i.e., God, YWH, Allah), and is given to us as a gift. Hence we are not allowed to reject it. From this vantage point, every (mental) crisis is a part of this gift and hence challenges us to throw our hopes on this powerful entity (i.e., crisis as a chance to demonstrate one's steadfastness of one's religious faith). This argument is adopted in Christianity by Paul (unknown, cir. 60 AC),

who denotes God's love as experienceable heaven on earth and as the ultimate shield against one's wish to enter eternal life immediately (ca. 54/55 BC, Philipper 1, 21–26), and in the Qur'an (Sura 4, 29). Basically, these arguments can be transferred into the *sanctity of life argument*, the classical *deontological argument* against suicide (see below).

(b) The *providence argument* claims that our lifespan and time of death (in this life) are laid out for us by a greater (transcendental) entity or power. This argument, already given in a certain sense in the Brahmanian and Hinduistic claim of bad karma caused by suicide, is put forward for Christianity by Paul and most explicitly by Thomas Aquinas (1225–1274) who argues that killing oneself usurps God's prerogative in determining when we shall die. This argument is also inherent in Allah's mercifulness toward Muslims: "O you who have believed, do not consume one another's wealth unjustly but only [in lawful] business by mutual consent. And do not kill yourselves [or one another]. Indeed, Allah is to you ever Merciful" (Qur'an Sura 4,29).

(c) Both arguments are present in antique Greek philosophy as well, highly influential at least for the Christian and, later, enlightened perspective. Plato (348/7–228/7 BC) argues in *Phaidon* that suicide is like deserting the flock without allowance of the herd. He draws on the property argument ("Yet I too believe that the gods are our guardians, and that we men are a possession of theirs"; Plato 2006, 62b) as well as on the providence argument ("Then, if we look at the matter thus, there may be reason in saying that a man should wait, and not take his own life until God summons him, as he is now summoning me"; cf. 62b, c) in order to justify his position. Unjustified suicide shall also imply bad conditions of the person's afterlife in the realm of the dead (*afterlife argument*). Basically, Plato's arguments against suicide are powerful due to his claimed reasonable acknowledgment of an immortality of the soul, allowing for the *afterlife argument*: A non-granted suicide will change the afterlife's conditions to the worst. This will be a powerful Christian argument as well: God will punish suicidents in their afterlife (see also: the cosmic order will bring about lesser reincarnations if you kill yourself in an unseemly manner).

(d) The *natural* or *cosmic law argument* claims that suicide stands up against the law as laid out for us in this reality/cosmos. This argument does not necessarily need a transcendental entity; it could refer to an evolutionary nature as well (natural law, *tian*). It is nonetheless a typical religious argument (see above). It is present in Augustine of Hippo's (354–430) claim that suicide violates the divine commandment "Thou shall not kill," since one kills a person (oneself). And it is explicitly given in Thomas Aquinas' second argument against suicide, claiming that suicide violates the order God established for the world. This argument is the fundament of both the property and providence argument, since both arguments are undergirded by the claim of a higher (divine) order established by this higher (divine) creator in the moment of creation (giving our life to us).

All three arguments were refuted by David Hume (1711–1776) during early Enlightenment, at least from a philosophical perspective. The *natural law argument* is circular reasoning, since "there is no being which possesses any power or faculty, that it receives not from its creator" (Hume 1995, #5). It could not be otherwise if we agree that "human life depends upon general laws of matter and motion," whether given by God or not (#4). Consequently, the *property argument* is difficult to adopt as well, since the received ability to kill oneself necessarily includes the liberty to dispose of one's earthly body, even if this body would belong to God (see below for the *sanctity of life argument*). Furthermore, the *providence argument* is not convincing, since every spontaneous action (including all kinds of inventions) must be in accord with providence; hence otherwise the office of providence would not be complete and continuous (#4–#5). Therefore suicidal behavior must be an integral part of the life laid out for the pertinent person. Hume ends up in debating the

problem of the "divine sign," which was the crucial challenge regarding all suicidal affairs for Plato and Paul, early Christians and Orthodox Jews. He argues that every hopelessness and desperation could be taken as such a "divine sign": "Whenever pain or sorrow so far overcome my patience, as to make me tired of life, I may conclude that I am recalled from my station in the clearest and most express terms" (#5). To summarize, the traditional arguments against suicide – including their ideas of exceptions – are both dependent on personal belief and circular reasoning (*petitio principii*).

Conceptual Clarifications/Definitions

Suicide and Parasuicide

Suicide means intentional self-killing (for old Greek: sui, self; cedere, to kill). Intentional self-killing can be performed by a great variety of behaviors (more "active" and more "passive" ones) but always involves actions by the relevant person displaying a direct and short-term connection between her intentional behavior and her death.

Suicide or better attempted suicide is often distinguished from so-called parasuicide. Parasuicide uses suicidal behavior as a gesture, urging others to help and rescue the suicidal person. Hence parasuicide is usually a "cry for help" (Stengel 1961). Suicidal behavior without intention to die can be lethal, too, although usually rather noneffective ("soft") methods of suicide are used compared to more harmful and effective ("hard") methods. Hence, from a suicide prevention perspective, all suicidal behavior should be taken seriously. There is ongoing debate from a suicide prevention perspective, whether this distinction is relevant or not.

Suicide and Self-Awareness

There is no suicide without self-awareness. In order to have the behavioral option to kill oneself, one needs to be aware of oneself in two ways: (a) consciousness about oneself and (b) consciousness or knowledge about one's mortality (Fairbairn 1995, p. 73f; Schlimme 2013). These reflective qualities of suicidal mental life do not imply that the suicidal person could not be able to kill herself without musing elaborately about her own mortality in that very moment. It is a well-known fact that people often report to have tried to kill themselves impulsively and without giving it elaborate consideration or having explicitly ambiguous second thoughts in the very moment of doing it. Yet, it is nevertheless a necessary condition to be aware of one's mortality in order to have this option of intentionally killing oneself, whether in an impulsive or an elaborately planned way. In other words, the option of suicide cannot be given without the reflective discovery of this behavioral option as a possible behavior for oneself. This usually takes place between the 10th and the 14th year of age and can, e.g., be lost in severe dementia.

Conceptual Clarifications of a Definition of Suicide

Against the first notion that the definition of suicide as intentional self-killing is rather precise, there is an ongoing debate whether such a definition is sufficient or not (Cholbi 2012). The concept of "intentional self-killing" does indeed require some clarifications:

(a) Suicidal behavior may not lead to one's death, but is suicidal behavior nonetheless because death was intended through one's intentional action. In this sense, suicide can be attempted but fails.
(b) Parasuicidal behavior may lead to one's death, but is parasuicidal behavior nonetheless because death was not intended through one's intentional action. In this sense, Fairbairn argues that suicide is "intentional rather than consequential in nature" (Fairbairn 1995, p. 58).
(c) If taken this way, "intentional" requires a non-coerced intention. In other words, "intentional" should not simply mean a self-consciously set-out goal for one's behavior *and* the decision which behavior is suited in order to achieve this goal, but it should imply that setting out this goal was a "free" decision of this person as well. If "free" means freedom from outer restrictions and coercions, i.e., by

other persons, then suicide would require the negative liberty to kill oneself if one wants to. If, however, "free" would mean freedom from inner restrictions and coercions, suicide would only be suicide if the decision to kill oneself is fully autonomous. It is highly questionable if the latter does even occur.

(d) However, psychological analysis of the suicidal state of mind (see below) reveals that suicidal persons are usually highly ambivalent regarding their intention to kill themselves *and* that their major motivation is not to die, but to escape from an unbearable mental condition (including their situational circumstances). If taking this point into account, suicide is not about intentional self-killing, but about intentionally escaping from an unbearable condition by the only way left from this person's point of view, that is, escaping from life. A meaningful definition of suicide would then be: Suicide means "intentional escaping from an unbearable life."

However, such a definition ignores that the suicidal person is well aware that her suicide may not necessarily lead to another life or the kind of life that such would be as well as the death which ensues from the suicidal act. It is indeed this fact that shapes suicide as a last rescue option. Furthermore, the escape intention is not already saying anything about concrete behavior. Consequently, a merely intentional definition of suicide is not sufficient. On the contrary, it is relevant that the intention of escaping from an unbearable life is *pursued* by the act of self-killing. Every definition of suicide must take into account the idea that suicide is basically about action.

As can be seen, a close conceptual analysis of suicide leads us to fundamental topics involved with the human behavior of suicide from a philosophical point of view. These topics concern freedom of the will (as regards setting out the goal of escaping unbearable life by means of killing oneself) and freedom of action (as regards the liberty to pursue one's set-out goal to escape unbearable life by means of killing oneself). Both topics culminate in the following questions:

Under which conditions is suicide morally justified?

Under which conditions is prevention of suicide morally justified?

Under which conditions is assisted suicide morally justified?

The Ethical Dimension of Suicide

The Suicidal State of Mind and the Medical Argument of Mental Illness

The term "suicidal state of mind" refers to those mental conditions in which a person thinks about killing herself and plans and prepares to commit suicide. To put it simple: People usually think about suicide when in a desperate state of mind. The suicidal experience is basically the experience of desperation *plus* the knowledge of suicide as one's last option to act in an effective way with respect to changing or altering one's feelings (one's desperation). Desperation, in the sense used here, does not simply mean emotions. It is, in fact, affecting all (active or passive) levels of mental life (i.e., both her pre-reflective valuations and perceived behavioral options as well as her reflective evaluations and behavioral possibilities; Schlimme 2013).

Especially during the last hundred years, the suicidal state of mind has been object of scientific investigation, mostly driven by suicide prevention intentions. A great variety of psychological theses regarding the suicidal state of mind has been proposed. While no single model may paint a sufficient picture of the suicidal state of mind, they nonetheless display more or less adequately how a suicidal person feels, why she is motivated to put herself to death, and why she probably performs suicidal behavior. The first methodological investigation of the suicidal state of mind was however performed by the Danish philosopher Søren Kierkegaard (1813–1855) (1992, p. 11f and p. 66ff). Although Kierkegaard was basically a religious thinker, he delivered exquisite descriptions of the desperate state of mind. He pointed out that the desperate person is well aware of her helplessness to cope with her desperation in all other regards (p. 12 and p. 66). This feature of suicidal mental life has

received different terms and concepts in psychiatric and psychological models (i.e., "narrowing," "hopelessness," "psychache"; Schlimme 2013).

While the desperation of suicidal persons has been highlighted in psychological models, there is nonetheless another side in the suicidal state of mind as well: the "hopeful quality" present in the option of escaping one's unbearable life through putting oneself to death. Kierkegaard was well aware of this attractiveness of one's death. Typically, positions claiming a right to kill oneself focus on this aspect of suicide (see below). Different manners of how suicidal people experience (value), or (prospectively) judge, their own suicide can be distinguished: as a "relief," a "remedy," or a "rescue" (Schlimme 2013).

The Janus-faced state of a suicidal mind corresponds to the oftentimes highly ambivalent interactions and behaviors of suicidal persons. This ambivalence seems moreover to be a stage in the suicidal process. According to Pöldinger, suicidal people usually live through at least two different stages before deciding whether to attempt suicide or not: (a) a stage in which the option of suicide is discovered and finally accepted as an option and (b) a stage in which the person constantly evaluates her life situation and behavioral options and seesaws between voting for suicide and further endurance of her seemingly unbearable life (Pöldinger 1982). The last stage is the phase after deciding to kill oneself, which usually corresponds with a sudden calming down of the suicidal person (like the quiet before the storm).

The medical perspective draws on the qualities of the suicidal state of mind, namely, the desperation of the suicidal person. The well-known facts that altered moods, delusional beliefs, and impaired cognitive abilities fuel desperation undergird the medical (psychiatric) argument that suicide comes about in phases of mental illnesses (*medical argument of mental illness*). And indeed, surveys demonstrate that roughly 9 out of 10 suicides are taken out while suffering from mental illness. A closer look reveals that this argument has two parts:

(a) The fact that mental illness implies passively suffered alterations of one's perceptions, evaluations, and thoughts. This influence challenges the notion that the intention to kill oneself is rational. The person might be influenced by a depressive mood, implying negativism (i.e., a non-intentional selection of negative attributes of oneself and negative/catastrophic outcomes of everything that can come to pass). Therefore the mentally ill person decides to suicide under coercion. She is not free due to "inner restrictions" impairing the rationality of her decision. This first part of the medical argument could be called the "argument of passivity." However, while this argument might be taken as a strong argument against the moral justifiability of suicide and has indeed been used in this rather general manner, it is relevant to realize that mental illness cannot in itself be an argument against the rationality of a given decision. On the contrary, it merely points out that the moral justifiability of a suicide decision is, at least from a modern or postmodern vantage point, dependent on the rationality of the concrete decision in the given situation.

(b) The fact that persons with mental illnesses, whether induced by substances (i.e., severe alcohol intoxication) or not, get well again and that others can assist and support recovery. Suicidal ideation during mental illness is typically a sign of a severe and critical illness episode and often lasts only for hours. During those episodes suicidal behavior can be performed impulsively, calling the decision's rationality into question. Typically, the suicide decision in this very moment during critical mental illness might be rejected by the same person only hours later. This temporary character of the suicidal crisis supports the medical argument of mental illness, claiming that the suicide decision might only hold temporarily and is hence not serving the person's best interests. This part of the argument could be called the "argument of temporariness." Both arguments are important to justify paternalistic interference (see below).

From a philosophical point of view, it is important to guard against arguments regarding the

three relevant moral questions drawn simply from the suicidal state of mind. Although the suicidal state of mind is an important venture point for arguing in a given case of intentional self-killing, it is nonetheless necessary to distinguish between the motivation for a certain intentional behavior and moral arguments regarding this behavior.

Actual Deontological Arguments Against Suicide

Drawing on the abovementioned three classical (religious) arguments against suicide, two deontological arguments remained after an enlightened critique: the sanctity of life argument and Kant's argument of a moral duty against oneself as ambassador of mankind.

The sanctity of life is an enlightened version of the property argument, since it does not rely on a divine entity or creator. A fundamental heteronomous quality of human life itself cannot be denied: Life is a given. From this fact of being given stems the notion of a "sanctity of life," entailing an inherent value of one's life and a call for reverence and respect from others and oneself for oneself. This implies that, since we are not able to give life to ourselves, we are not allowed to end our life. Consequently, suicide is forbidden since it negates the inherent value of one's life itself (*sanctity of life argument*). Since the fact of being given cannot be denied, this argument can be adopted from religious as well as nonreligious points of view. Nonetheless, it is typically connected with religious beliefs in divine entities (i.e., God), usually serving as givers of oneself (*property argument*). It can also be connected with the metaphysical idea of a "natural law" serving as a nonpersonal giver (*natural law argument*) and a "divine order" given by a creator in the first place (*providence argument*).

While the later arguments must face Hume's critique, there are also major difficulties with the sanctity of life argument (Cholbi 2012):

(a) Moral positions should be consistent. Since the sanctity of life must be granted to every human being, killing of another person is forbidden, even in case of self-defense. Even more problematic could be a position that claims sanctity of every kind of life (i.e., the principle of nonviolence *ahimsa* in Jainism), typically implying a thoroughgoing pacifism.

(b) The notion of an intrinsic value of life implies to keep alive even in the most horrifying situations (i.e., endless torture) or to endlessly sustain biologically alive bodies of persons with verified whole-brain death. If life is not taken as intrinsically valuable, this does not imply that life is valued as worthless. On the contrary, as Peter Singer argued the value of life is given extrinsically due to its intrinsic feature of being qualitative (Singer 1993). Quality of life is assessed from the first-person perspective regarding external features (i.e., future prospects, resources). It is an evaluation performed by each person herself. Singer's position argues (following the principle of maximizing well-being/happiness) that in case of an assessed low value of one's life, including negative future prospects, suicide may be justified. Taken that persons are interested in a high quality of life, this position claims that suicide may serve the person's best and dearest interest. However, assessments of quality of life are individual, hence highly controversial and difficult to define in a general manner (see below *authentic suicide*).

(c) Since living human beings are the embodied medium of the intrinsic value of life, it could be argued that suicide is not calling the intrinsic value of life into question, but that the relevant person judges other values to be more important (Dworkin 1993). Basically this position claims that the sanctity of life is a relative ("interpretive") moral principle, comparable to other values as well. Suicide decisions are hence never easy and often resemble a dilemma, implying that even an idealized judge could not come up with the one and only valid ("right") decision. This argument claims that life itself is not the highest value but can be compared with other values even though these other values are given for this person only in life itself (see below the Kantian argument against suicide). A similar position could be argued drawing on

Mengzi's argument cited above, since he rather compares two different ways of living with each other: a righteous versus a non-righteous way of living.

Another version of a deontological argument against suicide calls on our duties toward ourselves as human beings. From this vantage point, Immanuel Kant (1724–1804) argued that "the first duty of a human being against himself in his quality of animality is self-preservation in his animalic nature" (Kant 1968, p. 421). Kant is not claiming that this duty stems from some obscure drive of self-preservation. On the contrary, it is supposed to be a rational insight flowing from correct application of an inquiry of moral maxims (*categorical imperative*). For Kant it shall be a contradiction in terms (oxymoron) that the intention to kill oneself nullifies the fundament of this volition. Furthermore, Kant proposes a division of the person into a reasonable being (*homo noumenon*) and a bodily being (*homo phenomenon*), connected by a living trust relationship (*Treuhandverhältnis*) of the *homo noumenon* for the *homo phenomenon*. However, both points are not really convincing (Birnbacher 1990). If, however, the living trust relationship (*Treuhandverhältnis*) between reason (mind) and body would be taken as eternal, then a contradiction in terms would indeed be given. Kant himself proposed this way out as handwritten comments in a personal copy of his *Metaphysics of Morals* demonstrate (taking it as a regulative idea). However, with this turn Kant's argument is back in circular reasoning and subject to Hume's critique.

From a philosophical perspective deontological arguments are unable to principally argue a duty to stay alive under all conditions. Even if one accepts an intrinsic value of life, it is reasonable, at least from a philosophical point of view, to take it as only one value beside others, implying a moral (evaluative) conflict (dilemma) which indeed seems to be given in a suicidal state of mind. Furthermore, some philosophical positions argue that the moral evaluation does not compare life and death, but merely different ways of living even though one of these ways leads to one's death.

Social Arguments Regarding the Morality of Suicide

Human beings are living in families, communities, and societies. Insofar, each behavior has a social dimension. Consequently suicide might affect one's moral duties against others (i.e., loved ones, community) or against the state one lives in. Basically, two manners of relation can be given: Suicide can violate these moral duties *or* these moral duties can oblige one to suicide. Both manners of relation between a human being and its state have, for example, been discussed in ancient Greece. While the moral duty to suicide is usually embedded in the necessity to sacrifice one's life for the sake of the state, or for the sake of maintaining its order (cf. Greek tragedies), the first relationship is classically argued by Aristotle (384–322 BC). From his point of view, suicide is a premature and, therefore, irrational behavior "against rational (correct) reasoning" (Aristotle 2006, 1138 a9–14). Since the state's task is to promote rational behavior (from which a good life flows according to Aristotle), its laws forbid suicide.

Let us first consider social arguments of a duty to suicide. Especially military groups and armies might request potentially self-killing behavior from her members (suicide squad). However, it can be argued that life-risking behavior which intends to save or protect others should not be addressed as suicidal behavior (Cholbi 2012). Getting killed in action can therefore not be taken as suicide without an explicit statement of the pertinent person. Nonetheless, suicide might even be integrated in military tactics, as, for example, in ancient China General Sunzi (between 534 and 453 BC) did or as certain terrorist groups are doing at the beginning of the twenty-first century (i.e., suicide bombing combined with an eschatological interpretation of martyrdom). From a philosophical perspective, these examples point out the dilemma of external coercion. In a totalitarian social system, others can indeed be able to coerce someone into actually taking his life "voluntarily" (Battin 1994). Such a situation seemed to be given in Japanese culture during the shogunate periods (twelfth to nineteenth century). For certain (highborn, military) classes a highly

ritualized suicide, called *seppuku* in man and *jigai* in woman, was both a right AND a duty in certain situations in order to keep up their and their family's honor. Being part of their way of living, the pertinent persons prepared themselves for this last honorable act during lifetime and were indeed unable to suspend this supervised and well-prepared act (including hours, days, or weeks of mental preparation depending on the situation) in the case of given social obligation. The necessity for *seppuku* or *jigai* was typically given by command of one's lord (daimo) or cesar (shogun), resembling death penalties, or by loss of combat, possibly facing death anyway. However, deliberate and voluntary decisions on the basis of an internalized coercion due to growing up in a highly rigid social system permitting (or even demanding) suicide under certain conditions, drawing on soundly considered arguments and culture-specific values for this behavior, are difficult to reject as nonrational decisions from a philosophical perspective. Therefore, as a utilitarian argument could claim, it might be socially more prosperous to morally forbid suicide than to live in a society that permits suicide under certain conditions (Battin 1994). All debates on permitting suicide, euthanasia, and assisted suicide face this dilemma, usually put forward in the manner of a *slippery slope argument*.

Most philosophers argue that a moral obligation to kill oneself under certain conditions does not agree with the principle of *autonomy* which could be brought into play as *argument of self-property*. This is an important argument in the debate on euthanasia and physician-assisted suicide, since otherwise other persons could determine the moment of one being killed for certain reasons. Moreover, even a permission to suicide might especially affect vulnerable persons (adolescents, marginalized persons), resulting in a higher suicide rate in certain populations and raising the issue of equality and social justice. However, if suicide is permitted under certain conditions in a society, the pertinent person can expect *noninterference* from others (according to the moral principle of *autonomy*). She might also claim support from others in her suicidal behavior (according to the moral principle of *care* and *welfare*). Suicide assistance should be a voluntary act by the assistant and can, of course, be abused (Battin 1994). If a society intends to offer suicide assistance, it is indeed the crucial dilemma that it should be a best practice (i.e., regarding painlessness) and that practitioners must be kept from abusing and manipulating potential clients (Cholbi 2012).

Suicide might also harm one's moral duties toward others. Suicide causes distress, anguish, and a highly difficult bereavement process for suicide survivors (especially family members and loved ones). Besides mental harm, one's suicide can cause economic or material harm too (i.e., loss of financial support for dependents). From a strict utilitarian perspective, these harmful effects of suicide must be weighed against positive effects for the suicidal person (i.e., escaping an unbearable life without – subjective or objective – hope to change it to the better by any other means). Anyway, these social arguments can neither justify a moral obligation to stay alive nor to kill oneself. On the contrary, from this perspective suicide is morally an indifferent act whose moral evaluation depends on the actual situation and mental state of the suicidal person (i.e., socially isolated elderly male person with severe physical disabilities versus socially integrated mother of young children in excellent physical condition).

Social arguments against suicide might draw on one's obligation to contribute to one's community (society, state) welfare. Hume refuted this argument: "A man who retires from life does no harm to society: he only ceases to do good; which, if it is an injury, is of the lowest kind" (Hume 1995, #6). Nonetheless, one could argue that a community has a right to benefit from the special talents and labor force of each member, drawing on the reciprocal structure of the relationship between a person and her community (Cholbi 2012). However, as Hume argued the reciprocal structure is an obligation "to do good" to each other. Therefore even small benefits for a person outweigh small harms for her community (Hume 1995, #6–#8). Hume furthermore constructs ideal cases in which all social arguments against suicide fall away: "But suppose that it is no longer in my

power to promote the interest of the public; suppose that I am a burden to it; suppose that my life hinders some person from being much more useful to public" (#6). From this philosophical perspective, suicide would be permitted and could even be positively connoted, if one is socially "worthless" in the abovementioned sense and has already substantially contributed to one's community welfare. Again we face the slippery slope of an imagined culture that would cease to take care of the elderly, disabled, or marginalized, opening debates on equality, social justice, and social welfare (Birnbacher 1990; Battin 1994; Cholbi 2012).

From a philosophical point of view, social arguments can neither argue a duty to stay alive nor a duty to kill oneself. Two moral principles (namely, *autonomy* and *welfare*) are involved in the complex arguments weighed with each other. If we take into account that many suicides might lack both elaborate deliberation and support by sound arguments (i.e., fuelled by negativistically mistaken future prospects) – even though a period of evaluation took place – a (public) duty to prevent such irrational acts of others in one's community can be claimed (at least, if someone cares, as it should be the case in a community). While verbal interactions (i.e., calling on the person's good senses) are easily justifiable, physical restraint even for a short time (hours) requires sound arguments (i.e., medical argument of mental illness combined with social arguments drawing on the principle of *welfare*). On the other hand, deliberate suicide, supported by a rational line of argumentation, could at least claim noninterference or even assistance from others. Morally spoken, suicide is taken as an indifferent act whose moral quality is dependent on a variety of individual circumstances and conditions *and* the rationality of its justification.

Can Suicide Be Authentic?

If suicide as an indifferent act is morally justified according to the rationality of its justification, undergirded by its specific conditions, then it could be suspected that a fully rational suicide is possible. Rationality in the fullest sense would expect fixed conditions under which one's arguments are brought into play. However, two imponderables are given in the moment of suicide: (a) Since human beings can have no proven knowledge of death from a first-person perspective, comparisons between life and death are necessarily irrational, and (b) one's future prospects are not ultimately fixed. Suicide negates this openness, even if the remaining period of time is rather short (i.e., a person with a mortal physical illness) or a change of course is highly improbable (i.e., a person facing her death penalty after rejection of a death row pardon). Usually, however, the openness of one's future is far greater even if one suffers from unbearable mental illness. Consequently, a fully rational suicide is difficult to argue for, or justify.

Rationality of one's decision to suicide must hence stem from a rational comparison of two different manners of living (see above, Mengzi) under the imponderable condition of a principal future openness. On this basis five conditions can be defined as prerequisite of a rational decision: (a) an ability for causal and inferential reasoning, (b) a realistic (socially shared) worldview, (c) adequate information relevant for one's decision, (d) dying that enables one to avoid future harms, and (e) dying that accords with one's dearest and most fundamental interests (Battin 1994). While the first three conditions might be impaired due to mental illnesses, although this need not be the case, the latter two conditions are proposing a fundamental interest of human beings to live a good life. The avoidance of future harms is, however, a personal assessment open for influence from internal (i.e., a pre-reflective negativistic selection of future prospects due to one being severely depressed; Schlimme 2013) and external coercion (i.e., facing unjustified harm from others; see above). Given that all four conditions are met, the last condition is the crucial one and calls for an authentic decision (i.e., called for by the *Stoics*; see above, *Confucianism*). The claim that suicide can be an authentic decision has frequently been argued by philosophers in modernity (Birnbacher 1990; Battin 1994). However, since retrospective reevaluations are impossible, every suicide remains at least minimally insecure regarding its authenticity.

Conclusion

The moral evaluation of suicide is often problematic since different arguments, drawing on the moral principles of autonomy and welfare, must be weighed against each other. Impairments of this evaluation process, as can be given by mental illness ("inner coercion") and by external coercion (basically calling the classification of the behavior as suicide into question), may qualify the suicide decision as irrational. Such a nonrational suicide can expect to be prevented according to social arguments drawing on the moral principle of welfare with or without external restraint (strong versus weak paternalism). A fully rational suicide is principally impossible (justified belief is neither possible regarding one's future nor regarding one's condition after this life). Nonetheless, an authentic suicide displaying a rational justification (as rational as possible) can be given. Such a suicide can expect noninterference (strong paternalism) or even assistance (weak paternalism). However, since the arguments and values involved in the evaluation process are chiefly depending on the involved persons' assessments and dearest interests, the evaluation of a single suicide is individual and culture- and community-dependent. Hence evaluations can face a moral dilemma, implying that a "right" solution might not be found or not be agreed on in the relevant community.

Cross-References

- Assisted Suicide
- Autonomy
- Benefit and Harm
- Coercion
- Competence
- Death: Good Death
- Death: Meaning of
- Euthanasia: Active
- Euthanasia: International Debate
- Euthanasia: Passive
- Life: Meaning of
- Mental Illness: Concept of
- Paternalism
- Quality of Life
- Respect for Autonomy
- Right to Die
- Welfare

References

Aristotle. (2006). *Nikomachian Ethics. Nikomachische Ethik* (trans: .Wolf. Reinbek, U.). Germany: Rowohlt.

Battin, M. P. (1994). *The death debate. Ethical issues in suicide*. Upper Saddle River: Prentice-Hall.

Birnbacher, D. (1990). Selbstmord und Selbstmordverhütung aus ethischer Sicht. In A. Leist (Ed.), *Um Leben und Tod: Moralische Probleme bei Abtreibung, künstlicher Befruchtung, Euthanasie und Selbstmord* (pp. 395–422). Frankfurt/M: Suhrkamp.

Cholbi, M. (2012). Suicide. In E.N. Zalta (Ed.), Stanford encyclopedia of philosophy. http://www.plato/stanford.edu/entries/suicide/. Retrieved 19 Sept 2015.

Dworkin, R. (1993). *Life's dominion: An argument about abortion, euthanasia, and individual freedom*. New York: Knopf.

Fairbairn, G. J. (1995). *Contemplating suicide: The language and ethics of self-harm*. London: Routledge.

Hume, D. (1995). *Essays on suicide and the immortality of the soul: The complete unauthorized 1783 edition,* David Hume, *version 1.0,* James Fieser (Ed.),(Internet Release, 1995). http://www61.homepage.villanova.edu/kevin.hughes/Hume_Suicide.htm Retrieved 19 Sept 2015.

Kant, I. (1968). *Metaphysik der Sitten::Akademie Ausgabe*. Berlin: Walter de Gruyter.

Kiekegaard, S. (1849). *Sygdommen til Døden.* (1992). *Die Krankheit zum Tode*. (trans: Hirsch Gütersloh, E.). Germany: Gütersloher Verlagshaus.

Mengzi. Gaozi I. http://ctext.org/mengzi/gaozi-i Retrieved 19 Sept 2015.

Plato. (2006). *Phaidon*. [In Sämtliche Werke Vol. 2. Übers. F. Schleiermacher. Reinbek: Rowohlt.]

Pöldinger, W. (1982). Erkennung und Beurteilung von Suizidalität. In C. Reimer (Ed.), *Suizid: Ergebnisse und Therapie* (pp. 13–23). Heidelberg: Springer.

Schlimme, J. E. (2013). Sense of self-determination and the suicidal experience: A phenomenological approach. *Medicine, Health Care and Philosophy, 16*, 211–223.

Singer, P. (1993). *Practical ethics* (2nd ed.). Cambridge: Cambridge University Press.

Stengel, E. (1961). Selbstmord und Selbstmordversuch. In H. W. Gruhle, R. Jung, W. Mayer-Gross, & M. Müller (Eds.), *Psychiatrie der Gegenwart: Forschung und Praxis Vol. III Soziale und angewandte Psychiatrie* (pp. 51–74). Heidelberg: Springer.

Further Readings

Battin, M. P. (2003). Euthanasia and physician-assisted suicide. In H. LaFolette (Ed.), *The Oxford handbook of practical ethics* (pp. 673–704). Oxford: Oxford University Press.

Battin, M. P. (2010). The irony of supporting physician-assisted suicide: A personal account. *Medicine, Health Care and Philosophy, 13*, 403–411.
Shneidman, E. S. (1993). *Suicide as psychache: A clinical approach to self-destructive behavior*. Northvale: Jason Aronson.

Surgery: General

Peter Angelos
The University of Chicago, Chicago, IL, USA

Abstract

In recent decades, the ethical issues that arise in the care of surgical patients have received increasing attention. Although these issues in surgical ethics are not completely different from the ethical issues that arise in the care of nonsurgical patients, there are significant considerations that warrant specific attention in surgical patients. Central to the ethical issues in surgery are challenges to informed consent for surgery that requires surgeons to transmit information and engender their patients' trust prior to engaging in potentially dangerous acts. The necessity of weighing risks and benefits in surgery is clear since every surgical intervention carries the potential for significant harm to the patient. Innovations in surgery require specific attention since novel techniques designed to benefit specific patients do not require prior approval or review before a surgeon can undertake them. Surgical research, on the other hand, does require prior approval and specific research informed consent and also raises additional ethical considerations when compared to research in other areas of medicine. Although ethical issues in the care of surgical patients are not different in kind from the ethical issues in other areas of medical care, they are different enough in degree that they warrant specific attention.

Keywords

Surgery; Surgical ethics; Surgical innovation; Informed consent; Responsibility

Introduction

In the following paragraphs, important ethical issues that arise in the care of surgical patients will be examined. Although there are subspecialties in surgery that raise specific and often problematic issues (e.g., transplantation, pediatric surgery, and cosmetic plastic surgery), these areas will not be explored. Rather, focus will be placed on issues central to the care of all surgical patients. In particular, the following paragraphs will explore why, although surgery is a technical discipline, surgeons must be much more than technicians. Informed consent for surgery is central to the ethical care of surgical patients and will be examined. Informed consent requires not only a disclosure of risks, benefits, and alternatives, but also the patient must trust the surgeon to act in the patient's best interests even when the patient is completely vulnerable in the operating room. The responsibility that surgeons have for their interventions will be explored along with how that sense of responsibility shapes the surgeon-patient relationship. The inherent risks involved in surgery will be considered relative to the potential benefits. The necessity of weighing risks and benefits requires restrictions on the extent to which surgeons will go along with patients' wishes. The nature of surgical innovation will be examined relative to the differences with the process used to bring new drugs into practice. Although there is no requirement for outside review and oversight of surgeons' decisions to solve their patients' problems in novel ways, there are significant challenges for surgical research that will be explored.

History and Development

The ethical issues in the care of surgical patients are not different in kind but in degree when compared to the ethical issues in the care of any patient. The challenges of the doctor-patient relationship and the importance of communication are present for all clinical interactions between patients and their physicians. Similarly, the challenge of determining when treatments become too burdensome to patients is present in the field of

surgery as well as in nonsurgical fields of healthcare. There are no distinctively different standards for surrogate decision making in surgery compared to other areas of medical care. However, in the past few decades, it has become clear that in the broad category of clinical ethics, there are distinctive features of the ethical issues that arise in the care of surgical patients. Although no list of these distinctive features can be fully complete, in the following pages, attention will be focused on several of the most important items that influence most of surgical care of patients throughout the world.

In the early decades of the rise of medical ethics discussions, much of the work focused on patients cared for by internists, pediatricians, and obstetricians. There was a widely held view that surgeons were uninterested, and perhaps too busy, to take time to engage in extensive discussions about numerous alternative treatments for every patient. The old adage that "surgeons operate while internists discuss pros and cons" often suggested that surgeons were uninterested in nuanced discussions of ethical issues. However, in recent years, there has been a widespread acceptance by surgical organizations, academic surgeons, and community surgeons that the ethical issues in the care of surgical patients are significant and warrant attention. A few decades ago, many surgical residents and fellows would finish their training without ever having engaged in a formal discussion of the ethical issues associated with caring for patients. Today, most surgical trainees are acutely aware of many of the important ethical questions that arise in the care of their patients. These ethical issues that are central to the care of surgical patients will be considered in the following paragraphs.

Conceptual Clarification/Definition

The field of surgery is a very broad one with many different specialties and subspecialties that are all surgical. There are several areas of surgical practice that raise more ethical issues on a regular basis than others. Consider three surgical specialties that regularly raise numerous and specific ethical issues: transplant surgery, pediatric surgery, and cosmetic plastic surgery.

In transplant surgery, there are numerous issues involving an absolutely scarce resource (solid organs) and the many challenges of finding a just allocation system for that scarce resource. In addition, transplant surgeons must be attentive to the issues associated with both deceased donors and live donors. In the realm of deceased donors, questions of definitions of death become central along with questions about whether the deceased donor has any prerogative to direct the donation of his or her organs to a specific recipient or group of recipients. Certainly in the area of living donor transplantation, there are many ethical issues surrounding the questions of how to avoid coercive pressures on donors within families or among close friends, how to evaluate requests by people to be altruistic donors (i.e., to be live donors for unknown recipients), and whether donors should be paid for organs.

Within the area of pediatric surgery, numerous different ethical issues arise. How much discretion should parents have over deciding whether neonates should have lifesaving surgical interventions? When should doctors step in to assume guardianship of children whose parents decline to consent for curative operations? How should the well-being of an infant patient be safeguarded when the only possible treatment for a life-threatening disease or illness is a high-risk surgical procedure? These are just some of the many ethical issues that pediatric surgeons face on a regular basis.

In the realm of cosmetic plastic surgery, surgeons must consider what are the appropriate goals of medicine. For example, is there a standard of physical appearance that surgeons should try to uphold or should surgeons offer any body modification that the patient requests? Such a question becomes even more complicated when most cosmetic surgical procedures are paid for out of pocket by patients. A related question is how much risk a surgeon should be willing to subject his or her patients to when the patient is happy to assume the very high risks?

These are just a few of the many ethical questions that arise in the care of surgical patients in the areas of transplant surgery, pediatric surgery, and cosmetic plastic surgery. As important as the ethical questions in these specific areas of surgery are, in the following paragraphs, focus will not be on any of these specific issues. Instead, the focus will be on the more central ethical issues that all surgeons must address in their care of patients. Although surgery is clearly a technical discipline that demands merging detailed anatomic knowledge with operative skill that can only be obtained by thousands of hours of practice, surgeons are more than simply technicians. Every surgeon's encounter with a patient raises a series of important ethical issues that will be addressed in the upcoming pages. By focusing on the ethical issues central to the practice of any type of surgery, the considerations raised in the following paragraphs should have the most relevance to the largest numbers of surgeons and patients.

Ethical Dimension

Informed Consent

Central to the surgical care of patients is the requirement for surgeons to obtain informed consent from patients prior to undertaking any operative intervention on the patient. Although informed consent has not always been required prior to surgery, in recent decades, informed consent is a uniform requirement throughout the world. Although the legal requirements may vary among countries, the ethical requirement for informed consent prior to surgery is based on the importance of respecting the autonomous choices of patients. Surgeons are not allowed to operate on patients purely because the patient has a medical indication for surgery. Rather, a surgeon must explain to a patient why he or she needs an operation, as well as what the risks of the procedure are and what alternatives, if any, there are for the patient to consider. Only patients with the capacity to make autonomous choices are allowed to give informed consent after they have been told what are the risks, benefits, and alternatives to the surgery (Childers et al. 2009).

Although obtaining informed consent for an operation is a routine exercise for any surgeon, the process of explaining the operation to a patient and soliciting the patient's acceptance of the operation requires much more than simply the transmission of information. During the process of obtaining informed consent from a patient for surgery, the surgeon must encourage sufficient trust from the patient that the patient is willing to place his or her well-being in the surgeon's hands. Although this necessity for the patient to trust a physician is not unique to surgery, the level of vulnerability that surgery entails and the level of risk that the patient assumes are greater than in most other areas of medicine.

The Surgeon-Patient Relationship and Responsibility for Surgical Outcomes

The informed consent process is generally the beginning of the relationship between the surgeon and the patient. When giving informed consent to allow a surgeon to perform an operation, the patient is essentially acknowledging his or her willingness to trust the surgeon. Moving forward from that trust, the surgeon must assume responsibility for the patient's outcome. The relationship, therefore, between the surgeon and the patient is focused on the surgeon's commitment to act to benefit the patient. What is different between the surgeon-patient relationship and the more general relationship between physicians and patients is the surgeon's personal involvement in the surgery. Surgery does not "happen" the way, for example, a patient may respond to the administration of a medication. Surgeons "do" surgery on patients. In this manner, the surgeon's actions actually are the means for the treatment.

The nature of the surgeon's direct involvement in the patient's outcomes is perhaps best explained by Bosk who stated that "When the patient of an internist dies, colleagues ask, 'What happened?' When the patient of a surgeon dies, colleagues ask, 'What did you do?'"[Bosk 2003, p. 30]. Although this difference may appear to be a subtle one, it weighs heavily on surgeons in their

interactions with patients. Few surgeons would be willing to put a patient through a risky operation without a high degree of confidence that the patient will benefit from the operation (Kruser et al. 2015).

Inherent Challenge of Weighing Benefit and Harm

One of the central ethical issues in surgery involves the inherent challenge of weighing the risks and benefits of the operation. Every surgical procedure involves the very real possibility of harm. There is no surgical intervention that has only benefits. All operations begin with the deliberate actions of a surgeon that, in any other circumstance, would be considered harmful. For example, a straightforward surgical procedure such as removing a thyroid gland that has cancer in it involves first an incision. In any context other than in an operating room when a surgeon has the patient's informed consent, a slash of the knife across a person's neck would be a criminal activity. However, in the operating room when the surgeon makes an incision in the neck, this act is considered part of the therapeutic activity of the operation. Because every surgical procedure necessarily involves the potential for real harm, surgeons must carefully consider the possible benefits and whether they outweigh the risks.

As noted previously, informed consent for surgery requires the patient to be informed of the risks, benefits, and alternatives of an operation, and then the patient must choose to proceed with surgery. However, there may be circumstances when patients are willing to assume more risks than surgeons are willing to put them through. Although it is challenging to respect the autonomous choices of a patient while not abiding by those choices, surgeons are not ethically required to perform any operation that a patient wants regardless of the risks. Surgeons can recommend operations that patients may decline, but surgeons are not required to perform every operation that a patient may request. The autonomous choice of a patient extends to the reasonable medical or surgical options available but not to every possible option. In deciding whether to offer an operation to a patient, the surgeon must carefully weigh the risks of the procedure for the specific patient in view of the patient's medical condition and in view of the patient's goals.

The Nature of Surgical Innovation

One of the central differences between surgery and medicine emerges out of the differences between innovative drugs and innovative surgical procedures. In the USA, the Food and Drug Administration (FDA) must approve drugs that are offered to patients. Although the actual oversight may vary from country to country, every country requires new drugs to have at least some documented evidence for safety and efficacy before being used to treat patients. By contrast, there is *no* oversight of innovative surgical procedures. For example, when in the operating room and faced with complicated anatomy, surgeons are not only allowed to devise innovative surgical solutions to their patients' problems, but they are expected to seek such creative solutions. The ability to creatively solve a patient's problem in the operating room by the use of an innovative surgical technique is central to surgical practice and occurs without any regulation or oversight.

Innovative surgical techniques must be disclosed to patients preoperatively when planned before surgery or as soon as possible postoperatively if the innovation was unplanned (Biffl et al. 2008). Nevertheless, such innovative techniques need not have ever been tested prior to a surgeon trying it out on a patient. This level of creativity in surgery can be either beneficial or harmful to patients and is completely dependent on the surgeon's assessment of the risks and benefits to the patient. In this manner, even though patients must give informed consent prior to having any surgical procedure, patients must place their trust in their surgeons to make decisions about whether to use innovative techniques to benefit individual patients. The leeway that a surgeon is given to decide what procedure to perform on any given patient allows the surgeon a level of independence that is unparalleled in other areas of medicine.

When surgeons are obtaining informed consent from patients for planned innovative surgical procedures, by definition, the risks and benefits are not well known. In such circumstances, surgeons must carefully explain the uncertainty of

both the benefits and the risks. The disclosure of uncertainty is often difficult to explain since patients understandably expect that surgeons have knowledge of the procedures that are to be performed (Angelos 2010). The requirement for obtaining adequate informed consent for innovative surgical procedures is one that demands a level of transparency that is challenging for many surgeons and patients.

An additional challenge of innovative surgical procedures is that patients are completely dependent on the surgeon's assessment of whether the novel operation will be beneficial. By definition, there can be no well-documented evidence to support the use of a novel surgical procedure. Patients are forced to trust that their surgeons will accurately assess the risks and benefits even when these risks and benefits are not well known (Angelos 2014). If there were strong evidence to support the use of the innovative procedure, then it would no longer be truly innovative. When there is little data to determine if an innovative surgical procedure is safe, the patient must depend on the surgeon to provide a thoughtful assessment of risks and benefits and then accurately communicate these risks and benefits to the patient.

An important concept in the area of surgical innovation is that although novel techniques can be performed on patients with no oversight, surgeons must be careful not to plan innovative techniques for research purposes without obtaining prior approval from an institutional review board (IRB). There is a critical distinction between performing an innovative procedure on a patient *for the benefit of the patient* and performing the innovative procedure with the goal of *gathering the data* to answer the question of whether the innovative procedure is better or not (Biffl et al. 2008). The former circumstance is surgical innovation and is unregulated. The latter circumstance is surgical research and must therefore follow all regulations associated with human subject research.

Challenges of Surgical Research

As noted in the prior section, innovative surgical procedures that are performed for the benefit of the patient are not regulated. In other words, if a surgeon believes that he or she can help a patient by performing an operation in a novel manner, the surgeon can proceed to do that novel procedure without any prior review by anyone. In contrast, however, if a surgeon wants to study a novel operation to determine if it is better, then the surgeon is no longer engaged in innovative surgery *for the benefit of the patient* but rather is doing surgical research. The distinction between doing an innovative operation to benefit a patient and doing that same operation in order to gather data may seem like a subtle one. However, the intent of the surgeon is critically important. Although surgeons have tremendous latitude to creatively solve their patient's problems in innovative ways with no oversight, surgical research requires the same level of prior review and oversight that is required of all research involving human subjects.

In order to engage in research, surgeons must create a protocol that follows the same rigorous template as in any other medical research. In particular, surgical research must be carefully designed with a clearly written protocol. It must be based on prior research that justifies putting human subjects at any risk. Surgical research must be performed in a manner that minimizes risks to patients (who are also subjects), and there must be a formal informed consent document that outlines risks, benefits, and alternatives of participation in the research protocol. Furthermore, the informed consent document must have been reviewed and approved by an IRB prior to obtaining a subject's consent. The research consent form is different from the usual consent form for a therapeutic surgical procedure in that a research consent must spell out risks in much greater detail and should specify what is research and what is accepted therapy.

Even when a surgeon has gone through the steps of IRB approval for a surgical research protocol with all of the detail regarding inclusion and exclusion criteria and the subject/patient has signed the research consent form, there remains the challenge of variability in surgical practice. Medical research is focused on obtaining data to add to generalizable knowledge. The problem with surgical research is that there are myriad small ways in which surgeons differ in how they perform any given operation. In addition, there

may be many small variations in how the same surgeon performs the same operation in different patients. The anatomy of every patient is slightly different, and operations cannot follow a strict plan that ignores the differences of each patient. Although surgeons may, for the purposes of a research protocol, attempt to standardize an operation, there will be some level of variability that is inevitably present. This inherent variability does not make it impossible to perform surgical research but does create challenges for the ethical performance of such research.

Another challenge for surgical research involves randomization. The gold standard for medical research is the prospective randomized controlled trial. In trials of new medicines, the randomization involves the subject/patient being randomly assigned to receive the trial drug or standard therapy or sometimes the trial drug or a placebo medication. Often when surgical trials are randomized, a subject/patient will be randomly assigned to receive either the experimental surgical procedure or a standard operation. Perhaps it is not surprising, but many subjects/patients are uneasy about being randomly assigned to a specific surgical arm. For the reasons noted previously, surgeons are inherently involved in surgical interventions. The operation does not just happen, but a surgeon has to perform the operation. Perhaps for this reason, many subjects/patients are often very uneasy with the idea that the choice of their operation will be randomly assigned. This psychological finding does not make a surgical trial unethical, but it does push surgeons engaged in such trials to emphasize to subjects/patients that there is equipoise – namely, the surgeon does not know which treatment is better, and the study is designed to answer this question.

As noted in the paragraph above, placebos are often used in medication trials when there is no effective medical therapy. Placebo arms in surgical trials are much more problematic although not unethical (Angelos 2003). Often, when the outcomes of an operation are subjective, the potential placebo effect of actually undergoing an operation may be significant. In such cases, having a placebo (or sham) surgery arm may be critically important to determine if it is the actual surgical intervention that is the cause of the patient's subjective improvement, or rather they feel better because they have been through an operation and they assume that the surgery will help them. In order to distinguish between these possibilities, a placebo surgery arm can be very helpful. However, the more closely the sham operation resembles the "real" operation, the greater the risks for the subjects who are having the placebo surgery. For example, if the study is to determine if an operation on the abdomen alleviates a painful syndrome, then the sham procedure would need to involve more than just a general anesthetic. The sham operation would require an actual incision on the subject's abdomen that could become infected. Furthermore, if the "real" operation involves cutting through muscles to enter the abdomen, then in order for the sham operation to be more similar to the "real" operation, one might suggest that the muscles be cut in the sham operation as well. Thus, the closer the sham is to the "real" operation, the greater the risks to the subject. For this reason, a placebo surgical arm would carry risks to a research subject with no potential for benefit which makes placebo surgery essentially unlike a placebo drug or sugar pill.

Conclusion

In the previous paragraphs, several suggestions have been made regarding the ways in which surgical ethics is different from other areas of medical ethics. In particular, it has been shown that informed consent for surgery requires a high degree of trust by the patient in the surgeon. Informed consent for surgery, thus, requires the surgeon to communicate risks, benefits, and alternatives to patients so that competent patients can make autonomous choices. However, informed consent for surgery is more than simply the transmission of information. Surgical consent requires the patient to place great trust in the surgeon to perform potentially harmful and even disfiguring procedures that may ultimately benefit the patient. The need to constantly weigh the potential for harm against the potential for benefit shapes the manner

in which surgeons conceptualize their personal responsibility for their patients' outcomes.

In order for surgical care to improve in the future, there must be the possibility for change over time. Such change occurs in surgery most commonly as a result of individual surgeons' attempts to solve their patients' problems through creative approaches in the operating room. This unplanned surgical innovation is the primary driver for change over time. There is no requirement for oversight of this type of surgical innovation. In contrast to new drugs that require (at least in the USA) approval of the FDA, new procedures can be undertaken with no such formal review. This freedom to innovate in the operating room can potentially result in tremendous benefit or tremendous harm to patients. How surgeons manage such innovative procedures is central not only to the future of surgical care but also of the future of the surgeon-patient relationship. In order to prove that innovative surgical ideas are actually beneficial to patients, it is essential that there be well-constructed surgical research protocols. Although there are clear challenges to surgical research related to problems of standardization, randomization, and even whether placebo surgery arms can be used in clinical trials, surgical research is essential to provide the evidence necessary to ensure patient safety.

In the decades to come, more focused attention to ethical issues in the care of surgical patients will further enrich the overall understanding of clinical medical ethics. Through careful explorations of the similarities and differences between surgical ethics and medical ethics in general, surgeons will better understand how to best care for their patients in an ethical fashion.

Cross-References

- ▶ Bioethics: Clinical
- ▶ Bioethics: Medical
- ▶ Brain Death
- ▶ Clinical Equipoise
- ▶ Committees: Research Ethics Committees
- ▶ Consent: Informed
- ▶ Donation: Organs
- ▶ Face Transplantation
- ▶ Fetal Surgery
- ▶ Futility
- ▶ Organ Trade
- ▶ Organ Transplantation
- ▶ Placebo
- ▶ Professional-Patient Relationship
- ▶ Quality of Life
- ▶ Research: Clinical
- ▶ Respect for Autonomy
- ▶ Risk
- ▶ Surgery: General
- ▶ Transplantation Medicine

References

Angelos, P. (2003). Sham surgery in research: A surgeon's view. *American Journal of Bioethics, 3*(4), 65–66.

Angelos, P. (2010). The ethical challenges of surgical innovation for patient care. *Lancet, 376*, 1046–1047.

Angelos, P. (2014). Surgical ethics and the challenge of surgical innovation. *American Journal of Surgery, 208*(6), 881–885.

Biffl, W. L., Spain, D. A., Reitsma, A. M., Minter, R. M., Upperman, J., Wilson, M., Adams, R., Goldman, E. B., Angelos, P., Krummel, T., & Greenfield, L. J. (2008). Responsible development and application of surgical innovations: A position statement of the Society of University Surgeons. *Journal of the American College of Surgeons, 206*, 1204–1209.

Bosk, C. L. (2003). *Forgive and remember: Managing medical failure* (2nd ed.). Chicago: University of Chicago Press.

Childers, R., Lipsett, P. A., & Pawlik, T. M. (2009). Informed consent and the surgeon. *Journal of the American College of Surgeons, 208*(4), 627–634.

Kruser, J. M., Pecanac, K. E., Cooper, Z., Steffens, N. M., McKneally, M. F., & Schwarze, M. L. (2015). "And I think we can fix it": Mental models in high-risk surgical decision making. *Annals of Surgery, 261*(4), 678–684.

Further Readings

Grant, S. B., Modi, P. K., & Singer, E. A. (2014). Futility and the care of surgical patients: Ethical dilemmas. *World Journal of Surgery, 38*(7), 1631–1637.

Namm, J. P., Siegler, M., Brander, C., Kim, T. Y., Lowe, C., & Angelos, P. (2014). History and evolution of surgical ethics: John Gregory to the twenty-first century. *World Journal of Surgery, 38*(7), 1568–1573.

Pugliese, O. T., Solari, J. L., & Ferreres, A. R. (2014). The extent of surgical patients' understanding. *World Journal of Surgery, 38*(7), 1605–1609.

Sade, R. M. (2015). *The ethics of surgery: Conflicts and controversies*. New York: Oxford University Press.

Surrogacy

Shamima Parvin Lasker
Department of Anatomy MH Samorita Medical College, Dhaka, Dhaka, Bangladesh
American University of Sovereign Nations, Scottsdale, AZ, USA
Bangladesh Bioethics Society, Dhaka, Bangladesh
Asian Bioethics Association, Bangkok, Thailand

Abstract

Surrogacy is a promising treatment for infertility. It can potentially solve many intolerable difficulties that the infertile couples and their families face. Although initially frowned upon, evidence shows that the surrogacy arrangements are more acceptable now than when it was first introduced. Therefore, changes in the attitude in decision making about surrogacy can also be seen in some countries, but there are still indications of the degree of divergence between discourse and the actual practice of different forms of surrogacy around the world. Social, ethical and legal problems are subject to major debates and disagreements in natural or partial surrogacy or genetically unrelated full surrogacy. Genetic gestation surrogacy may largely free from social, legal and moral complications. It is a great choice of infertility treatment if the couple want their own genetic baby, but it still requires more thoughts and discussion. This chapter attempts to discuss the different notions related to surrogacy worldwide.

Keywords

Surrogacy; Decision-making

Introduction

Infertility affects 15 % of reproductive couples globally; around 50–80 million people worldwide may experience infertility. It is believed that approximately 10 % of global infertility occurs in developing countries. People in sub-Saharan Africa alone experience three times more infertility than other regions (WHO 2010).

Since the birth of the first test tube baby, Louise Brown in 1978, assisted reproductive technology (ART) has evoked great interest amongst the public. ART including egg or sperm donation or a surrogacy arrangement encourages infertile couples, giving a means of immense hope. But this new type of treatment for infertility has created legal and ethical debate among different societies and the followers of different religions all over the world. Major debate, disagreement, and controversy have arisen relating to natural or partial surrogacy or genetically unrelated full surrogacy. Genetic gestational surrogacy (when the sperm and ovum of a husband and wife/couple is fertilized by an in vitro fertilization [IVF] technique and the embryos are transferred to a surrogate host) is largely free of social, legal, and moral complications. It is a great choice of infertility treatment if a couple want their own genetic baby (Brinsden 2003). Nevertheless, the widespread use of such technologies is prohibited by some cultures.

Although the problem of infertility has not been totally eliminated, over the years scientists have been able to come up with better ways to help childless couples to fulfill their dreams of having a child. However, there are issues regarding the notion of surrogacy in different societies and cultures; these are evaluated in this chapter.

Social Conditions of Infertile People

There is a very specific relationship between parents and children. There is an unconditional, firm belief that this newborn baby will always be important and lovable to parents – no matter what happens. It is believed that a child enriches the parents' life, and that quality of life will deteriorate when the desire for parenthood is not fulfilled – non-fulfillment can make people very unhappy. As a consequence, infertile couples experience great social distress and face an intolerable situation, especially those living in more

traditional and conservative societies or who are from lower socioeconomic classes.

In some societies, childless couples are excluded from taking part in leading and important roles in family functions such as birthdays, weddings, and other events involving children. Infertile people are viewed in some countries as a burden to the socioeconomic well-being of a community due to a loss of continuity. In these communities, children confer social status, guarantee rights of property and inheritance, and provide continuity by maintaining the family name (Ombelet et al. 2008). Lack of a child in some cultures is considered a lack of old-age security. In these cultures, the child assists with labor and provides a reciprocal duty to parents in their old age (Lasker 2011). In some countries, having children is a social obligation, that is owed to the husband's family (Abbasi-Shavazi et al. 2008). In egalitarian societies, people want children as part of their life plan and they suffer when they cannot fulfill this wish. Because parenthood has deep social roots, the social and psychological consequences of involuntary childlessness are often severe and have a large impact on people's lives.

Although male infertility has been found to be the cause of failure to conceive in about 50 % of cases, the social burden falls disproportionately on women. In some societies, the social status of the women, her dignity, and self-esteem are closely related to her procreative potential in the family and society as a whole. Childbirth and child rearing are regarded as family commitments and not just biological and social functions.

In some cultures, infertile women often live in fear that their marriage will collapse. In fact, in some cultures personal status laws consider a wife's barrenness to be a major grounds for divorce, e.g., Islam. Islam also allows women to divorce if male infertility is proven. Morally, women usually do not take initiative of divorce on the grounds of infertility unless their marriage is truly unbearable (Inhorn 2006). In some cases, the husband or his family consider a second marriage as a solution. This second marriage, however, may be a great misfortune for the first wife and permits the man to be polygamous (WHO 2010). Childless women are frequently stigmatized, resulting in isolation, neglect, and domestic violence (Ombelet et al. 2008), all of which violate human dignity. Consequently, it is not surprising that many infertile couples may do whatever it takes to save their relationship, even if it means undertaking risky or expensive treatment. Due to the lack of insurance coverage for infertility treatments in some countries, these costs are heavy and even unbearable to poor people. Therefore, prevention and treatment of infertility are of particular significance around the world.

Definition and Classification of Surrogacy

Surrogate literally means "substitute." The terms "surrogacy" or "surrogate mother" are usually applied to the woman who carries and delivers a child on behalf of another couple. When the intended surrogate is inseminated with the semen of the husband of the couple, the procedure is known as straight surrogacy or traditional surrogacy. Gestational surrogacy is when a sperm and an ovum from a couple is fertilized by IVF and transferred to a surrogate; this is also called genetic gestational surrogacy. When a sperm or an ovum is donated by a third party, fertilized by IVF, and transferred to a surrogate, it is called gestational partial surrogacy. If both the sperm and ovum are donated by a third party, fertilized by IVF, and the resulting embryo is transferred to a surrogate, it is known as gestational full surrogacy. In gestational surrogacy, the surrogate mother is not genetically related and will be free from all responsibilities after delivery of the child (Brinsden 2003).

Surrogacy can be either commercial or altruistic. When the surrogate is paid for donating the egg/sperm or for gestation of the fetus, or both, it is called commercial surrogacy. If the surrogate is unpaid, it is regarded as an altruistic surrogacy.

Indications for Surrogacy

The indications for surrogacy are congenital absence of the ovaries/testes/uterus; men with

azoospermia; women who have had a hysterectomy for carcinoma or hemorrhage but who still have functioning ovaries; women who have suffered repeated miscarriages and for whom the chance of ever carrying a baby to term is remote; women who repeatedly fail to implant a normal healthy embryo in their uterus; or certain medical conditions, such as cancer or heart or renal disease, which might threaten the life of a woman (Brinsden 2003).

When couples are incapable of producing ova/sperm as a result of disease (e.g., cancer), injury or normal aging, a donor ovum/sperm may be fertilized in vitro and implanted in a surrogate's uterus, and they then gestate the baby to term. The couple may choose this type of surrogacy with the hope that the child will be at least half-related to them.

Surrogacy opens the way for post-menopausal women or women once considered hopelessly barren to have a child despite having no genetic link to it. This practice is also an option for single people or homosextual couples who wish to have a child and can enable women who may not want to become pregnant because of their busy schedule to become a mother (Zawawi 2012). However, it is important to make clear that the obsessional and symbolic desire to have a child by surrogacy arrangement should not morally be encouraged due to ethical issues discussed in this chapter.

Historical Background of Surrogacy

Before the advent of modern ART, straight surrogacy (also known as traditional or natural surrogacy) was the only means of helping childless women to have a child, and it has been practiced since ancient times. One of the oldest examples of surrogacy is the story of Abraham, Sarah, and Hagar in the Bible. Sarah, Abraham's wife, was barren. In order to maintain his lineage, Abraham went to Hagar, a maid, who he later married. Hagar give birth a son, Ishmael. Sarah became jealous of Hagar and Hagar did not want to give the baby up to Sarah to raise with Abraham (Genesis 16). Another example of surrogacy in the Bible comes from Jacob and his wives, Rachel and Leah. While Leah gave birth to four sons, Rachel remained barren. She became jealous of Leah and gave Jacob her maidservant, Bilhah, to be a surrogate mother for her. Bilhah gave birth to two sons: Dan and Naphtali (Genesis 30:3). Another Biblical precedent for surrogacy is Mosaic Law, which provided for levirate marriage (a type of marriage in which the brother of a deceased man is obliged to marry his brother's widow, and the widow is obliged to marry her deceased husband's brother), an example of which was when Boaz, family member, impregnated his dead brother's widow Ruth to bear children on his behalf (Genesis 38).

Examples of surrogacy are also found in the ancient Indian scripture, the Mahabharata. According to the Mahabharata, Gandhari, the wife of king Dhritarashtra, conceived and the pregnancy went on for nearly 2 years, after which she delivered a mass (mole). Lord Vyasa found that there were 101 cells that were normal in the mass. These cells were put in a nutrient medium and were grown in vitro to full term. Of these, 100 developed into male children and one into a female child. Hindu mythology presents many more examples of surrogacy.

The Code of Hammurabi (1780 BC) indicates the presence of surrogacy 1800 years before the birth of Christ, and it is likely that it was the first legal document relating to traditional surrogacy arrangements. The Code regulated and controlled the legal grounds of surrogacy, and was mainly used to advocate producing male offspring in Mesopotamia (Svitnev 2006). Surrogacy was also quite common in ancient Egypt – many pharaohs used their concubines to produce male heirs. However, even though the children delivered by these maids were treated as the pharaoh's children, their rights were somewhat reduced. They could assume the throne only if there were no other nobler and more legitimate contenders. Traditional surrogacy was also common in ancient Greece and Rome (Svitnev 2006).

Before the advent of ART, natural surrogacy was the only means of helping childless women to have children. Later, artificial insemination became a more acceptable means of achieving pregnancy than natural surrogacy. Now,

surrogacy by IVF has become a successful treatment. In late 1976, the first reported baby, known as Baby M, was born by gestational surrogacy in the USA. In 1980, the first commercial surrogacy arrangement was made in the USA, with Elizabeth Kane being paid US$10,000 to act as a traditional surrogate. Later, she became an advocate against surrogacy, and wrote a book entitled *Birth Mother* expressing her experiences of emotional difficulties with children, family, and society. Currently, there are only two sources of very rough statistics on surrogacy, and these report numbers relating to gestational surrogacy only. Based on available statistics data from the US Centers for Disease Control and Prevention (CDC), there were a total of 45,870 live births from gestational surrogacy in 441 clinics in the USA during 2009. In addition, several thousand more babies are born each year as the result of a wide variety of surrogate arrangements worldwide. The numbers of surrogacies in Australia, Canada, and Brazil are at least as large as those reported in the USA.

Surrogacy Around the World

Surrogacy is a promising treatment for infertile couples. When IVF is not possible, surrogacy may be an alternative choice for many couples. However, different opinions relating to surrogacy exist around the world. Some examples of this are discussed below.

Both partial gestational and commercial surrogacy are allowed in the USA and Canada. However, different US states have different regulations: ten states (Arkansas, Florida, Illinois, Nevada, New Hampshire, North Dakota, Texas, Utah, Virginia, and Washington) have laws allowing surrogacy under certain circumstances; seven states (New York, Michigan, Arizona, Nebraska, Kentucky, New Jersey, and Indiana) and Washington, DC, have laws that prohibit, penalize, or void surrogacy contracts. California is one of the most permissive surrogacy states, although there is no legislation relating to the practice (Perez 2010). Surrogacy is not allowed in South America.

Like the USA, Australia has different regulations in different states. In New South Wales, Western Australia, and the Australia Capital Territory, surrogacy is freely available. Surrogacy is not illegal in Victoria, South Australia, and Tasmania, but the strict controls regulating surrogacy and payments relating to it make it almost impossible to be carried out commercial surrogacy in these regions. But the altruistic surrogacy is legal throughout Australia.

Diversity is apparent within the 27 member states of the European Union. In Austria, Germany, France, Italy, and Switzerland, surrogacy is prohibited. Criminal sanctions are applied for non-compliance, ranging from heavy fines to imprisonment. According to German legislators, surrogacy should be prohibited because of the violation of bonus mores (morality). The UK, Belgium, The Netherlands, and Finland are the only countries in Europe Union that allow surrogacy. In the UK, patients can be treated by gestational surrogacy for exceptional reasons after intensive investigation and counseling. Commercial surrogacy arrangements are illegal. Surrogacy UK, and COTS (Childlessness Overcome Through Surrogacy) are charitable and non-profit organizations involved in surrogacy in Britain. Treatment cannot take place outside the legal cover provided by the Human Fertilization and Embryology Act 1990 in the UK (Brinsden 2003). In Spain, surrogacy contracts are null and void but surrogacy per se is not prohibited. Spain is the European epicenter of reproductive tourism. Spanish egg donation is often done altruistically by Spanish women with or without monetary compensation (Inhorn et al. 2010).

Initially, Italy had developed one of the most cutting-edge ART industries in the world, earning the moniker of "the wild west" of assisted reproduction in Europe. However, in 2004 the Italian parliament banned all types of reproductive technologies (including contraception, abortion, IVF, third-party gamete donation, and surrogacy). The resulting Medically Assisted Reproduction Law is known as Law 40/2004. The moral justifications given for this new law are (1) the possibility of incest; (2) lineage; (3) problems with biological paternity; and (4) the risk of positive eugenics

(creating a child with sought-after characteristics of a donor, e.g., blue eyes, blonde hair, IQ >130) (Inhorn et al. 2010).

In Greece, gestational surrogacy has only been allowed since 2002. If the commissioning mother is married, the written consent of her husband is required, and the intended patients must provide a medical attestation of her inability to gestate the child. In addition, both the prospective parent and the surrogate mother must reside in Greece (Svitnev 2006).

Commercial surrogacy is legal in most of the countries of the former Soviet Union (e.g., Armenia, Belarus, Georgia, Kazakhstan, Kirgizia, the Russian Federation, and Ukraine). Russia is considered as a sort of reproductive paradise. The Basic Law of the Russian Federation for Citizen's Health Protection states that each adult woman of childbearing age has the right to artificial fertilization and the implantation of an embryo (Svitnev 2006). There is no concept of the right to fatherhood in Russia, but single men applying for surrogacy to become fathers should be treated equally in accordance with the equal rights and freedoms of men as citizens, regardless of sex. Written informed consent of all parties is required for participation in the surrogacy procedure. Apart from that consent, neither adoption nor a court decision is required. Russia is also one of the very few countries in which posthumous surrogacy can be arranged.

Israel legalized surrogate motherhood in 1996. The surrogate can be paid only for legal and insurance expenses and can be compensated for her time, loss of income, and pain.

All type of surrogacy is allowed in South Africa. However, a child born as a result of an invalid agreement is deemed to be the child of the woman who gave birth to that child. A surrogate mother who is also a genetic parent of the child may terminate the surrogate motherhood agreement at any time by filing a written notice with the court (Svitnev 2006).

Though surrogacy in Asia is a gray area, a 2009 report by Reuters estimated that around 25,000 children have been born in China by means of commercial surrogacy arrangements. The "womb-for-rent" industry defies the country's strict childbirth laws. Reuters added that three young surrogate mothers were discovered by authorities in Guangzhou and forced to abort their fetuses (Svitnev 2006). In India, where commercial surrogacy has been popular since 1992, surrogacy is not yet directly mentioned in law. However, only Indian citizens aged 21–35 years can become surrogates. Korea operates ART without statute or guidelines (Svitnev 2006).

All of the Muslim countries except Iran and Lebanon issued bioethical decrees in 1980; these support assisted reproduction treatments but disapprove all types of third-party ART. Gamete donation and surrogacy are prohibited for three major reasons: (1) adultery; (2) the potential for incest among the offspring of unknown donors; and (3) genealogical lineage. The prohibition of gamete donation and surrogacy has been enacted either by law or by professional medical codes of ethics in 62 Muslim countries throughout the world, e.g., Egypt, Sudan, Morocco, sub-Saharan Muslim countries, Kuwait, Qatar, Saudi Arabia, the United Arab Emirates, Bahrain, Syria, Turkey, Indonesia, Malaysia, Afghanistan, Pakistan, India, Bangladesh, etc. (Inhorn 2006). However in 1999, the Supreme Jurisprudent of the Shiite branch of Islam, Ayatollah Ali Hussein Khamanei, the handpicked successor to Iran's Ayatollah Khomeini, issued a verdict that permits donor technologies including surrogacy. This ruling is gaining acceptance within some of the Shiite population in Iran and part of Iraq, Lebanon, Bahrain, Syria, Saudi Arabia, Afghanistan, Pakistan, and India, etc. (Abbasi-Shavazi et al. 2008).

Surrogacy and Religion

With globalization, researchers, doctors, and patients alike are moving around to different parts of the world. Thus, it is becoming common that physicians may have to provide medical services to patients with ethical precepts that are different from their own. Physicians need to be sensitive to this diversity and avoid a stereotyped approach to religious patients. Healthcare professionals and researchers should be aware of

different religious backgrounds relating to surrogacy that may help clinicians and researchers to better understand and negotiate the dynamics of each physician–patient relationship before they make a judgment regarding medical practice. Therefore, we provide here a short review of the main religious traditions in the world and their general attitudes towards surrogacy.

The Catholic Church is strongly against all forms of assisted conception, particularly those associated with gamete donation and surrogacy (Zoloth and Henning 2010). However, among Catholic believers there are wide varieties of attitudes and ideas about surrogacy, showing a complex reality that varies in closeness to the Vatican's teachings. The Anglican Church is less rigid in its views and has not condemned the practice of surrogacy.

The value of procreation is depicted in Genesis 1:28 where God's first command to human beings is to "be fruitful and multiply." Therefore, third-party donation of gametes, including surrogacy, is allowed in the Jewish religion, which sees procreation as a duty of the Jewish man to have children. In the Jewish religion the child born as a result of surrogacy will belong to the father who gave the sperm and to the woman who gave birth (Schenker 1997, p. 113). According to Jewish Halakhic law, single Jewish women are preferred as a surrogate, both to avoid the implications of adultery for married surrogate women and to confer Jewishness, as Jewishness is seen to be conferred through the mother's side, particularly through the act of gestating and birthing the baby. However, most conservative rabbis prefer that non-Jewish donor sperm should be used in order to prevent adultery between a Jewish man and a Jewish woman and to prevent future genetic incest among the offspring of anonymous donors (Inhorn 2006).

Traditional Hindu literature, especially the Mahabharata, depicts Kunti, Madri, and Gandhari – three queens to ensure that there will be children and the Bharata family lineage will continue. Hindu bioethics not only permits but strongly encourages using ART to have a child, particularly when a couple has had difficulty conceiving and especially to have a son (Bhattacharyya 2006). The Mahabharata considers non-genetic and genetic children as morally and meaningfully equivalent. Therefore, ideas of family extend beyond the nuclear family of parents and children to include aunts, uncles, in-laws, adoptive relatives, grandparents, close friends, and even all of the members of the town in which an individual was raised. Since lineage does not depend on a genetic tie between parents and children, children need not be genetically related to their fathers to count as heirs in Hindu culture (Bhattacharyya 2006).

Some Buddhist schools encourage or at least accept ART because it aims to alleviate future suffering as a result of infertility. Therefore, Buddhists accept all types of ART including surrogacy as long as the technology brings benefits to the couple who wish to have a child and it does not bring pain or suffering to any parties involved. Buddhism would find no conflict in applying and using modern technology. However, some Buddhist schools criticize ART for perpetuating the disillusioned attachment to life that sometimes motivates human beings to sensual desires. Although ART may remove the physical and bodily desires of sex from the reproductive process, the mental or emotional aspiration of the couple, child, or third party can be problematic. Some monastic texts, such as the Vinaya Ptaka, equate the desire for a child with the desire for wealth and economic security that leads humans astray from the path to enlightenment. In addition the Dhammapada declares that "one's body belongs to oneself or one's child belongs to oneself." A non-genetically related child can no more belong to a parent than a genetically related child. Some Buddhist thinkers may, therefore, eschew ART for exacerbating disillusioned notions about the parent–child relationship that might, arguably, be harmful to both parent and child (Numrich 2009).

Muslims are divided into two main schools of thought: Sunni and Shiite. The majority (90 %) of Muslims throughout the world are Sunni. In 1980, Sunni scholars permitted treatment using all types of ART but disallowed any form of third-party reproductive assistance, including surrogacy. Use of third-party gamete donation for

reproduction is problematic as it violates the precepts of Islam concerning legitimacy, lineage, inheritance, and incest (Inhorn 2006). Another problem resulting from surrogate motherhood is who is the real mother? In the Qur'an, the definition of motherhood is that mothers are those who conceive and give birth to the baby (*walada hum*). The Arabic verb *walada hum* is used for the whole process of begetting, from conception to delivery. It does not only refer to the act of carrying (*haml*) and giving birth (*wad*'). Thus, this Qur'anic verse categorically denies any rights of motherhood to the genetic mother. Determining who the mother is in the case of genetic gestational surrogacy is a problem (Kabir and az-Zubair 2007).

Conversely, Shiite scholars do allow surrogate motherhood as a treatment for infertility, albeit only for legal couples. Shiite scholars consider the embryo to be different from sperm and so do not regard introducing the embryo into the womb of the surrogate mother to be the same as introducing the sperm of a man to whom she is not married. In fact, they regard the surrogacy arrangement as transferring an embryo or fetus from one womb to another and do not see any sin in this practice (Abbasi-Shavazi et al. 2008).

Conclusion

The available literature shows that surrogacy decisions are based on the moral, religious, and philosophical principles of the society in which they are undertaken. Furthermore, the ethical and social implications are deeply intertwined with religious traditions and communities. Which countries prohibit surrogacy depends on what religion the majority of the population belongs to and what the precept on procreation is of that particular religion. The literature shows that artificial reproductive technology is allowed in every country but all types of third-party assisted reproduction is restricted in some countries and cultures on the basis of adultery, preservation of lineage, inheritance, potential incest among the half-siblings, and possible eugenics. Major debate, disagreement, controversy, and ethical and legal problems have arisen from natural or partial surrogacy or genetically unrelated full surrogacy. Genetic gestation surrogacy may bypass these problems. In a genetic gestational surrogacy arrangement, there is no chance of incest among the half-siblings as the genetic relationship is already known, and there is no fear of confusion of lineage of the child as the biological parents are already confirmed. In this arrangement, the surrogate mother does not actually engage in any act of adultery, as it does not involve any body contact of a sexual/adulterous nature; therefore, the punishment of adultery is not applicable. This is because it is belived by some groups that introducing a third party is presumably problematic as it introduces a third party into the sacred dyad of husband and wife relationship that may threaten the marital bond. Motherhood may be problematic in genetic gestational surrogacy in some cultures; however, since neither the biological mother nor the surrogate has comprehensively fulfilled the definition of motherhood, according that culture, motherhood can be confer to genetic mother by weighing the public benefit and necessity within the marriage bond. As genetic gestational surrogacy is largely free of social, legal, and moral complications, it can be used to provide the highest form of happiness to couples for whom the concept of family was previously impossible.

Competing Interests

The author declares that she has no competing interests. The author conceived the idea of the article and wrote the manuscript.

References

Abbasi-Shavazi, M. J., Inhorn, M. C., Razeghi-Nasrabad, H. B., & Toloo, G. (2008). The Iranian ART revolution: Infertility, assisted reproductive technology, and third-party donation in the Islamic Republic of Iran. *Journal of Middle East Women's Studies, 4*(1), 1–28.

Bhattacharyya, S. (2006). *Magical progeny, modern technology: A Hindu bioethics of assisted reproductive technology.* Albany: State University of New York Press.

Brinsden, R. P. (2003). Gestational surrogacy. *Human Reproduction Update, 9*(5), 483–491.

Inhorn, M. C. (2006). Making Muslim babies: IVF and gamete donation in Sunni versus Shi'a Islam. *Culture, Medicine and Psychiatry, 30*, 427–450.
Inhorn, M. C., Patrizio, P., & Serour, G. I. (2010). Third-party reproductive assistance around the Mediterranean: Comparing Sunni Egypt, Catholic Italy and multisectarian Lebanon. *Reproductive Biomedicine Online, 21*, 848–853.
Kabir, M., & az-Zubair, B. (2007). Who is a parent? Parenthood in Islamic ethics. *Journal of Medical Ethics, 33*, 605–609.
Lasker, S. P. (2011). Infertility treatment in developing country. *Bangladesh Journal of Bioethics, 1*(2), 3.
Numrich, P. D. (2009). The problem with sex according to Buddhism. *Dialog, 48*(1), 62–73.
Ombelet, W. I. C., Silke, D., Serour, G. I., & Devroey, P. (2008). Infertility and the provision of infertility medical services in developing countries. *Human Reproduction Update, 14*(6), 605–621.
Perez M. (2010). Surrogacy: the next frontier for reproductive justice. http://rhrealitycheck.org/article/2010/02/23/surrogacy-next-frontier-reproductive-justice/. Accessed 25 Nov 2015.
Schenker, J. G. (1997). Infertility evaluation and treatment according to Jewish law. *European Journal of Obstetrics & Gynecology, 71*, 113–121.
Svitnev K. (2006). Legal control of surrogacy – International perspectives. pp. 149–63. http://www.jurconsult.ru/publications/ethical_dillemas/13_Legal%20control%20of%20surrogacy%20-%20international%20perspectives.pdf. Accessed 27 Aug 2015.
WHO (World Health Organization). (2010). Mother or nothing: The agony of infertility. *WHO Bulletin*, 88(12), 881–882. http://www.who.int/reproductivehealth/publications/infertility/en/index.html. Accessed 23 Mar 2014.
Zawawi M. (2012). Third party involvement in the reproductive process: Comparative aspects of the legal and ethical approaches to surrogacy. Eubios Ethics Institute. http://www.eubios.info/ABC4/abc4389.htm. Accessed 1 May 2015.
Zoloth, L., & Henning, A. A. (2010). Bioethics and oncofertility: Arguments and insights from religious traditions. *Cancer Treatment and Research, 156*, 261–278.

Further Readings

Fasouliotis, S. J., & Schenker, J. G. (1999). Social aspects of assisted reproduction. *Human Reproductive Update, 5*(1), 26–39.
Nachtigall, R. D. (2006). International disparities in access to infertility services. *Fertility and Sterility, 85*(4), 871–875.
Norhayati, H. A. (2003). Assisted reproduction – Islamic views on the science of procreation. *Eubios Journal of Asian and International Bioethics, 13*, 59–61.
Utian, W. H., Sheean, L., & Goldfarb, J. M. (1985). Successful pregnancy after in vitro fertilization-embryo transfer from an infertile woman to a surrogate. *New England Journal of Medicine, 313*, 1351–1352.

Sustainability

Darlene Fozard Weaver
Duquesne University, Pittsburgh, PA, USA

Abstract

Various meanings of sustainability point to complex, interlocking challenges for global bioethics. In global bioethics, sustainability can refer to conditions necessary for maintaining particular health-care systems; to economic development, particularly for underserved populations; and to health care that does not damage the natural environment. This entry reviews the historical development of sustainability, the tensions between economic and environmental emphases in sustainability, and the import of sustainability for global bioethics. Sustainability comprises a norm that increasingly informs global bioethics, though the term itself can obscure thorny ethical issues.

Keywords

Sustainable development; Environmental bioethics; Social determinants of health

Introduction

In health care, the term sustainability can refer to the capacity to sustain programs or systems that provide preventative care or treatment. The terminology of sustainability also appears prominently in literature on economic development, public health, and in literature on human health and environmental issues. Economic and environmental sustainability are urgent, interconnected issues in global bioethics. They bear on and complicate the sustainability of health-care programs and systems, for example, affecting access to antiretroviral drug treatment for HIV/AIDS in resource poor contexts or straining already financially burdened health-

care sectors when climate-related natural disasters occur. Sustainable health care promotes human health without doing ecological damage and in concert with larger strategies for improving the lives of poor and marginalized persons in a globally interdependent context. Sustainability comprises a norm that increasingly orients and informs work in global bioethics. However, the meaning of sustainability is itself disputed in ways that highlight complex moral questions at the intersection of human health, economic development, and care for the environment.

Meanings of Sustainability

There is no single definition of sustainability in global bioethics. The term itself can refer simply to the continuance of a given health-care initiative or the viability of a particular system or organization, and thus operate as a relatively nonmoral term. More often sustainability functions normatively. In global bioethics, the concept is often drawn from development discourse and environmental ethics. Sustainability can operate in moral discourse that is concerned primarily with development, referring to the eradication of poverty and promotion of economic growth. This principally economic version of sustainability can exhibit varying degrees of attention to and concern for the natural environment. Sustainability can also refer principally to environmental protection, conservation, and repair. Here, too, the concept permits a wide range of viewpoints regarding the intrinsic value of the natural environment, the moral status of nonhuman species, and attitudes toward trade-offs among human health, economic growth, and well-being of the natural environment. The concept of sustainability often joins economic development to environmental concerns (at least in the form of stewarding natural resources for future generations). One notable and often used example comes from *Our Common Future*, also known as the Brundtland Report, which was sponsored by the United Nations World Commission on Environment and Development (UNCED). The Brundtland Report defined sustainability as "meeting the needs of the present generation without compromising the ability of future generations to meet their needs" (UNCED 1987). While the Brundtland Report as a whole does speak to morally relevant questions like whether the natural environment has intrinsic value, its definition of sustainability leaves such questions open. Following the Brundtland Report, the concept of sustainability in development discourse usually jointly refers to a goal of economic development that avoids, minimizes, or seeks to repair environmental harm (du Pisani 2006).

The Brundtland Report identifies three pillars of sustainability: environmental protection, economic development, and social equity. These dimensions of sustainability intersect but do not overlap. Tensions arise between economic growth, for example, and environmental protection. The relationship between health and human development typically falls under the pillar of social equity. More recently, discussions of sustainable development have come to include a fourth pillar of sustainability, that of culture (UNESCO 2001). Appeals to culture resist the reduction of sustainable development to material benchmarks by explicitly acknowledging the importance of cultural diversity and intangible cultural heritage. Designating culture as a pillar of sustainability affirms the indispensable role culture plays in providing a coherent basis for human participation in the development of material sustainability and motivation to support the changes that are necessary for that development (UNESCO 2001).

A variety of allied concepts appear in lieu of or in concert with sustainability, including but not limited to conservation, resilience, human security, and stewardship. With regard to environmental sustainability, conservation refers to the protection and preservation of natural resources, environments, and populations. It sometimes includes the repair of ecosystems or wildlife populations. Resilience increasingly appears in discussions of climate change; it refers to the capacity to undergo stress yet maintain function. Resilience can be a feature of natural and built environments, as well as ecological and anthropogenic organizations or systems. Resilience is also

used to refer to capacities to adapt to changing conditions (Pierce and Jameton 2003). The concept of resilience increasingly appears in development discourse (Sustainable Development Solutions Network 2014). The concept of human security situates human health and well-being in a framework that includes attention to violence and peace, economic security, and environmental security (Lautensach and Lautensach 2015). It fosters an integrated, multidisciplinary approach to security risks and to paradigms of well-being (Lautensach and Lautensach 2015). Another concept allied with sustainability is stewardship (Resnik 2012). It appears in secular as well as religious literature on topics as diverse as biomedical research and land management. While it, too, is a contested idea, stewardship generally refers to the prudent management of resources entrusted to one's care. It can complement the principle of sustainability by directing attention to the particular responsibilities of diverse moral agents.

Sustainability in Bioethics: A Brief History

The genesis of environmental bioethics is often attributed to Van Rensselaer Potter's *Bioethics: Bridge to the Future*, in which he argued for a broad understanding of bioethics (Potter 1971). Rather than limit it to medical quandaries faced by individuals, Potter envisioned bioethics as a new philosophy that bridged disciplinary divides among medical, biological, and environmental sciences and between the sciences and humanistic inquiry. Globally, the environmental movement was already gaining traction. Rachel Carson's *Silent Spring*, published almost a decade earlier than Potter's *Bioethics*, raised awareness regarding the impact of pesticides and other toxins (Carson 1962). Garret Hardin and others voiced concerns about population growth (Hardin 1968). Indeed, Potter's vision for bioethics as a field concerned with the survival of the species was part of a larger global emergence of environmental awareness and advocacy. Nevertheless, bioethics developed principally as medical ethics, focused primarily on individual patients and their best interests. Potter later called his approach to bioethics "global bioethics" to further distinguish its scope and orientation from medical ethics (Potter 1988).

As recognition of environmental degradation grew in the 1970s and 1980s, so did awareness that high rates of economic growth and technological progress do not prevent economic disparity. Moreover, this disparity threatened prospects for continued rates of economic growth. Efforts to promote development in impoverished regions appeared at odds with environmental conservation, which seemed to limit acceptable forms of development (du Pisani 2006). In his historical study of the concept of sustainable development, Jacobus A. du Pisani argues that it emerged as a compromise between growth and conservation; the concept of sustainable development brought significant elements from previous development discourse – notably themes of progress and growth – into relation with environmental degradation and growing economic inequality (du Pisani 2006).

In the 1980s, the concept of sustainability increasingly came to refer to human development undertaken in an environmentally mindful fashion. The Brundtland Report's understanding of sustainability, with its pillar of environmental protection, provided a basis for a number of high-profile global initiatives, including the 1991 United Nations Environmental Program report, *Caring for the Earth*, the 1992 United Nations Conference on Environment and Development (UNCED), the emergence of Agenda 21 from UNCED, the Earth Charter in 2000, the World Summits on Social Development in 2002 and 2012, the UN Millennium Development Goals, and many others. With the 2015 deadline for the Millennium Development Goals at hand, work is under way to update targets and indicators for Sustainable Development Goals (Sustainable Development Solutions Network 2014). As the following section will show, however, the increasing use of sustainability as a concept belies serious disagreements of its meaning and scope.

While the concept of sustainability gained traction in development discourse, it also became more prominent in bioethics through engagement

with economic and environmental issues. Scholarship focused on poverty and social determinants of health connected bioethics to public health and to development discourse. Paul Farmer's *Pathologies of Power* revealed shortcomings in some development strategies, which can be imperialistic, by undertaking tuberculosis treatment in Haiti in a manner informed by liberation theology (Farmer 2005). Norman Daniels developed a framework for allocation of health-care resources that is indebted to John Rawls' *Theory of Justice* (Daniels 2008).

Others have promoted bioethical concern for sustainability through research on the public health impact of environmental degradation and, conversely, the impact of health-care systems on the environment. Jessica Pierce, for example, advocates for environmentally sustainable practices in medical and hospital settings and for environmentally friendly health-care products. Pierce and Andrew Jameton coauthored *The Ethics of Environmentally Responsible Health-Care*, which not only explored the impact of environmental degradation on human health and health-care but also argued that aspects of medical practice contribute to the environmental harm that undermines human health (Pierce and Jameton 2004). More recently, David Resnik published *Environmental Health Ethics*, the first English monograph devoted to environmental health (Resnik 2012). Resnik argues that environmental health ethics needs to supplement the traditional principles of health ethics with principles that expand the scope of our moral concern beyond humans to include other species and the natural environment. He identifies six principles from environmental ethics for this task: utility, environmental justice, animal welfare, stewardship, sustainability, and precaution.

A growing concern for sustainability is evident in the variety of global initiatives, coalitions, and academic programs that marry health-care, development, and ecology. For example, the Earth Summit in Rio de Janeiro (1992) yielded Agenda 21, a voluntary program to promote sustainable development. In 2012, the United Nations Conference on Sustainable Development reaffirmed commitments to Agenda 21. Sustainable coalitions among health systems are emerging; the 2020 Healthcare Climate Challenge, for example, is a worldwide campaign intended to mobilize hospitals and other health-care systems to reduce their adverse environmental impact and protect public health from climate change. Academic programs facilitate collaboration among scholars, governmental leaders, and nongovernmental capacity builders. Examples include the United Nations University, the UNESCO/UNITWIN (University Twinning and Networking) Networks Programme, as well as degree programs at colleges and universities around the world.

While there is growing evidence that bioethics is seriously engaging issues of sustainability, bioethics focused on environmental issues and global economic inequality remain subfields or niche interests. Bioethicists who are committed to doing bioethics in an environmental and economically sustainable fashion continue to call their colleagues to conversion, urging them to consider the participatory role bioethics can play in fostering sustainable public health programs and policies, "green" medical therapies, and sustainable research protocols (Lautensach and Lautensach 2015; Resnik 2012; Pierce and Jameton 2003; Valles 2015). Will global bioethics continue to integrate economic and environmental dimensions of sustainability? Will bioethical engagement with poverty and ecology remain subfields, or will they transform bioethics more broadly?

Sustainability as a Moral Norm

Research on social determinants of health shows an undeniable correlation among poverty, other forms of social vulnerability, and poor health outcomes. Poverty is linked with higher rates of chronic disease, poor nutrition, inadequate access to clean water, and premature death, to name a few (Sustainable Development Network Solutions 2014; Daniels 2008; Farmer 2005). The maldistribution of vulnerability, access to social and material goods, and opportunity through social structures, policies, and cultural processes amounts to a structural form of violence against persons (Farmer 2005). Improving health

outcomes requires nothing short of dismantling these structures. Sustainable development is often suggested as a solution to health disparities, though it is important to note that approaches to sustainable development can themselves contribute to rather than ameliorate structural violence (Farmer 2005).

Damage to the environment imperils human health in a variety of ways, including but not limited to exposure to toxins and pollutions, depleted natural resources, food insecurity, climate-related and vulnerability to natural disasters, and vulnerability to communicable diseases (Pierce and Jameton 2003; UNCED 1987). Environmental sustainability appears vitally important for human health (Sustainable Development Solutions Network 2014). Yet it is also important to note that environmental sustainability can conflict with the promotion of public health. For example, the World Health Organization has endorsed the limited use of DDT (dichlorodiphenyltrichloroethane) to combat malaria even though DDT use can cause environmental damage (Carson 1962; Resnik 2012). Granting that conflicts such as this will inevitably arise, sustainability in its economic and environmental dimensions designates a normative orientation to understand human health in its manifold relations to ecological and social systems and to improve health by addressing these systems.

Sustainability and Contested Moral Questions

The meaning and force of sustainability as a moral norm for global bioethics depend significantly on several interrelated and contested moral questions. To begin, there is the question of what moral obligations we have to future generations. Using the Brundtland Report's definition of sustainability, our obligation is to avoid compromising the ability of future generations to meet their needs. Understanding the scope of this obligation would require bioethicists to parse the term "needs," particularly in relation to wants, and to consider different ways of conceptualizing quality of life. Potter identifies five modes of life (mere, miserable, unjust, idealistic, and acceptable) that are relevant to such inquiry (Potter 1988). Even once an obligation to future generations is clarified, there remains a question regarding its force vis-à-vis moral obligations to currently existing human persons and to the natural environment. On what grounds do we adjudicate conflicts between the needs of future generations and the needs of present human lives? Does sustainability place limits on certain medical endeavors, for example, weighing against the use of resources for nontherapeutic cosmetic surgery, in the interest of curtailing their environmental impact and reallocating those medical resources to programs that serve vulnerable populations? What is the import of sustainability for end of life care? Daniels develops an account of health as a special social good – one essential for pursuing and enjoying a host of other goods – in terms of normal species functioning; his theory of justice, indebted to John Rawls, endeavors to show when health inequalities are unjust and when limits on health are fair (Daniels 2008). Some limits on health can then be shown to be warranted in order to protect future generations' equality of opportunity for health as normal species functioning (Daniels 2008). Under a strong formulation of sustainability, regard for future generations is an urgent moral question with the potential to overturn the received wisdom of bioethics on even the most basic matters. More modestly, sustainability can inform particular judgments regarding ordinary and extraordinary care.

In any case, the question of humanity's debt to future generations is complicated by the massive economic inequality found within particular regions and globally. Research on social determinants of health clearly shows the interrelationship of poverty and poor health outcomes. Discussions of sustainability are variously optimistic or pessimistic regarding the compatibility of economic growth and environmental conservation. Development discussions of sustainability soberly note the scale of global economic inequality and the urgency of our ecological crisis, yet articulate development goals predicated on the ultimate feasibility of promoting both economic growth and environmental sustainability (Sustainable Development Solutions Network 2014). Lautensach and Lautensach, however, consider the fact that human beings currently consume natural resources at the rate of approximately 1.4 planets

and that overconsumption permanently damages the biosphere thereby reducing the Earth's capacity for supporting future generations; they argue that the terminology of sustainable growth is simply nonsensical from a scientific standpoint (Lautensach and Lautensach 2015).

A second contested question, related to the first, concerns what we owe to nonhuman species. While this issue includes matters such as the use of nonhuman animals in biomedical research, its import is much larger. As Lautensach and Lautensach note, as many as several dozen nonhuman species disappear every day (Lautensach and Lautensach 2015). The cumulative extinction of so many species impacts the biosphere in which human beings are only one species along many others, threatening the very conditions necessary to sustain human life. The loss of agricultural biodiversity includes threats to food systems and nutrition, endangers resources for traditional medicinal treatments and pharmacological research for new therapies, and increases risks for infectious disease.

Bioethics can make consequentialist appeals to human self-interest to raise concern for biodiversity. However, instrumentalizing nonhuman species or planetary biodiversity may simply replicate anthropogenic moral patterns. So a third contested question is whether the natural environment has intrinsic worth. If so, what constraints does it place on particular human choices? Disagreements over the precautionary principle can illustrate differences regarding the intrinsic worth of ecosystems, nonhuman species, or biodiversity. The precautionary principle often figures in discussions of sustainability and global bioethics. The precautionary principle warrants a course of action, such as protective or preventative measures in order to avoid harm, even in the absence of scientific consensus about the risk of such harm (Resnik 2012; Jordan and O'Riordan 2004). The precautionary principle has been invoked, for example, to support a ban on genetically modified organisms (GMOs) as well as hydraulic fracturing for natural gas, although its application in both cases is challenged by supporters of these technologies.

As these contested questions indicate the normative force of sustainability depends greatly on the sort of moral commitments one builds into its meaning, such as commitments to social equity or distributive justice, the environment, and future generations. Even when a particular medical practice or health system is "green," or environmentally sustainable, its overall moral quality could be problematic. A sustainable practice can be morally wrong on other grounds. Therefore, other norms must work in concert with sustainability. Norms like beneficence, nonmaleficence, autonomy, and solidarity can complement sustainability. If sustainability is not integrated with other bioethical norms, environmental bioethics and global bioethics will remain niche subdisciplines within bioethics more broadly. If growing calls for more sustainable bioethics are heeded and sustainability becomes a central norm, it has potential to transform bioethics.

Sustainability and Other Norms

Given this history, when sustainability is used normatively, it typically signals a commitment to social equity, to distributive justice, and to environmental protection, conservation, and repair where the latter is possible. Because appeals to sustainability in bioethics, environmental ethics, and development discourse entail the contested questions discussed above, and because the relative force of the concept's economic and environmental dimensions is ambiguous, the application of sustainability in bioethics requires additional normative judgments. As a moral norm, sustainability must work in concert with other norms from bioethics, economic ethics, and environmental ethics (Resnik 2012). It is also important to note that a particular medical practice or system for allocating health-care resources could be sustainable yet still be morally problematic on other grounds. Sustainable health care is not equivalent to morally acceptable health-care.

Sustainability, Health Care Industry, and Research

Bioethicists are increasingly documenting the impact the health-care industry has on the environment (Pierce and Jameton 2003, Richie).

Environmental impact includes energy use, pollution, hazardous chemicals, and waste management, along with participation in unsustainable services, for example, through overuse of disposable products and other purchasing practices. Sustainability, however, may require changes that go well beyond reducing the adverse impact of health-care systems. Once again, the question arises whether an emphasis on the environmental component of sustainability has the normative force to overturn, or at least seriously challenge, ostensibly settled moral convictions in bioethics. A controversial article that explores the idea of what an environmentally sustainable reproductive technology looks like provides an example (Richie 2014). All human births, whether or not they are facilitated by assisted reproductive technologies (ART), contribute to population growth and humanity's carbon footprint. Cristina Richie, however, argues that the impact of fertility clinics on the environment differs morally from the impact that occurs through natural reproduction, since ART is undertaken as a commercial medical therapy. She argues that fertility clinics should be subject to carbon caps and that ART funding should be eliminated for those who are not biologically infertile. The argument is merely one example of the way commitments to sustainability can chafe against other moral commitments, in this case to reproductive liberty.

Sustainability also bears normatively on research and research ethics (Resnik 2012). With regard to economic development and human health, commitments to sustainable development include building capacity for research ethics committees in contexts where human subjects have been exploited (Resnik 2012). Commitments to sustainability in global bioethics also underscore the importance of research aimed at alleviating poverty, understanding the impact of environmental degradation and climate change on public health, and devising sustainable therapies and health-care systems (Pierce and Jameton 2003). However, a concern for sustainability increases ethical tensions regarding the funding of research projects. Pressure to accept funding from biotechnology and pharmaceutical corporations in the absence of robust governmental funding can create conflicts of interest in any case. Bioethicists could be more reluctant to speak against the adverse environmental impact of particular corporations or their products.

Bioethics and Advocacy

By its nature as a normative enterprise, bioethics seeks to impact the practice of medicine as well as the health systems in which medical practice is situated. This impact need not rise to the level of advocacy for social or structural change. A norm of sustainability, however, even in its weaker versions, beckons bioethicists to intervene against climate skepticism; to advocate for greener practices and infrastructures; to collaborate with a wider array of scientific, governmental, and nonprofit partners; and to engage in public intellectual activity necessary to build support for sustainable policies and practices. Increased advocacy by bioethics can benefit from the work of health behavior psychologists and communications experts (Valles 2015). Research on scientific communication suggests that framing climate change risks in terms of public health risks better elicits emotional responses that correlate with support for action to mitigate climate change (Valles 2015). Collaboration with experts from a wide variety of sectors is a necessary component for morally responsible as well as effective advocacy. Consider a hypothetical example of such advocacy in which bioethicists urge for legislation to provide access to air conditioning for vulnerable populations to lower their risk of heat-related death during climate change-induced heat waves (Valles 2015). While this advocacy could be warranted by appeals to sustainability, a commitment to sustainability would also require bioethicists to understand the ecological footprint of air conditioning, including its impact on energy systems and ecosystems. Moreover, that impact assessment would need to morally evaluate the program in light of a debt to future generations. To the extent that the environmental dimension of sustainability is emphasized, bioethicists will need to expand the scope of their interlocutors as well as their range of vision. For some bioethicists, this expansion must include challenges to the very feasibility of global

development goals that are premised on the legacy of development discourse from the Brundtland Report onward (Lautensach and Lautensach 2015).

Conclusion

Economic and environmental factors are crucially important determinants of human health. They also impact the health-care systems that are necessary for preventative care and treatment. Worsening economic inequality and the impact of climate change represent urgent public health challenges for health-care systems. Sustainability, in the threefold sense discussed here, comprises a positive norm for global bioethics, albeit one fraught with internal tensions. A normative commitment to sustainability challenges contemporary bioethics to develop new frameworks for bioethics, to undertake more interdisciplinary and cross-cultural collaboration, and to measure expenditures of human, social, economic, and environmental capital in terms of their impact on vulnerable human populations and our imperiled environment.

Cross-References

- ▶ Advocacy
- ▶ Agricultural Ethics
- ▶ Animal Ethics
- ▶ Animal Rights
- ▶ Animal Welfare
- ▶ Anthropocentrism
- ▶ Biocentrism
- ▶ Biodiversity
- ▶ Bioethics: And Politics
- ▶ Bioethics: Environmental
- ▶ Biopolitics
- ▶ Biosecurity
- ▶ Climate Change and Health
- ▶ Development and Bioethics
- ▶ Development Ethics
- ▶ Disasters
- ▶ Disease
- ▶ Equality and Equity
- ▶ Food Ethics
- ▶ Food Security
- ▶ Future Generations
- ▶ Genetic Modification (GMOs): Animals
- ▶ Genetic Modification (GMOs): Food
- ▶ Genetically Modified Organisms (GMOs): Human Beings
- ▶ Global Ethics
- ▶ Health: Global
- ▶ Human Rights
- ▶ Justice: Global
- ▶ Pandemics
- ▶ Poverty
- ▶ Precautionary Principle
- ▶ Public Health
- ▶ Quality of Life
- ▶ Religion and Global Bioethics
- ▶ Solidarity
- ▶ Stewardship
- ▶ UNESCO

References

Carson, R. (1962). *Silent spring*. Boston: Houghton Mifflin.

Daniels, N. (2008). *Just health: Meeting health needs fairly*. Cambridge: Cambridge University Press.

Du Pisani, J. A. (2006). Sustainable development: Historical roots of the concept. *Environmental Sciences, 3*(2), 83–96.

Farmer, P. (2005). *Pathologies of power: Health, human rights, and the new war on the poor*. Berkeley: University of California Press.

Hardin, G. (1968). The tragedy of the commons. *Science, 162*, 1243–1248.

Jordan, A., & O'Riordan, T. (2004). The precautionary principle: A legal and policy history. In M. Martuzzi & J. A. Tickner (Eds.), *The precautionary principle: Protecting public health, the environment and the future of our children*. Copenhagen: World Health Organization.

Lautensach, S. W., & Lautensach, A. K. (2015). Environmental security: International, national, and human. In D. Moellendorf & H. Widdows (Eds.), *Routledge handbook of global ethics*. New York: Routledge.

Pierce, J., & Jameton, A. (2003). *The ethics of environmentally responsible health care*. Oxford: Oxford University Press.

Potter, V. R. (1971). *Bioethics: Bridge to the future Englewood*. Cliffs: Prentice Hall.

Potter, V. R. (1988). *Global bioethics: Building on the leopold legacy*. East Lansing: Michigan State University Press.

Resnik, D. (2012). *Environmental health ethics*. Cambridge: Cambridge University Press.

Richie, C. (2014). What would an environmentally sustainability reproductive technology look like. *Journal of Medical Ethics*. doi:10.1136/medethics-2013-101716. Available: http://jme.bmj.com/content/early/2014/05/20/medethics-2013-101716.short. Last accessed 2 Aug 2015.

Sustainable Development Solutions Network. (2014). Health in the framework of sustainable development: Technical report for the post-2015 development agenda. Available: http://unsdsn.org/wp-content/uploads/2014/02/Health-For-All-Report.pdf. Last accessed 2 Aug 2015.

UNESCO. (2001). Universal declaration of cultural diversity. Available: http://portal.unesco.org/en/ev.php-URL_ID=13179&URL_DO=DO_TOPIC&URL_SECTION=201.html. Last accessed 17 June 2015.

United Nations World Commission on Environment and Development (UNCED). (1987). Report of the world commission on environment and development: Our common future. Available: http://www.un-documents.net/wced-ocf.htm. Last accessed 17 June 2015.

Valles, S. A. (2015). Bioethics and the framing of climate change's health risks. *Bioethics, 29*(5), 334–341.

Further Readings

Dobson, A. (1998). *Justice and the environment: Conceptions of environmental sustainability*. Oxford: Oxford University Press.

Green, R. M., Donovan, A., & Jauss, S. A. (2008). *Global bioethics: Issues of conscience for the twenty-first century*. Oxford: Oxford University Press.

Hattingh, J. (2014). Protection of the environment, the biosphere, and biodiversity. In H. ten Have & B. Gordijn (Eds.), *Handbook of global bioethics*. Dordrecht: Springer.

Holman, S. R. (2014). *Beholden: Religion, global health, and human rights*. Oxford: Oxford University Press.

Synthetic Biology

Michael J. Selgelid[1] and Nicholas Evans[2]
[1]Centre for Human Bioethics; School of Philosophical, Historical and International Studies, Monash University, Clayton, VIC, Australia
[2]Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Synthetic biology is a techno-scientific discipline with the declared goal of rationally engineering biological systems. Despite its considerable promise – regarding applications in medicine, energy, environmental remediation, and agriculture – synthetic biology raises numerous ethical issues pertaining to intellectual property, the creation of novel life forms, biosafety, and biosecurity.

Keywords

Synthetic biology; Genomics; Intellectual property; Genetically modified organisms; Justice; Biosafety; Biosecurity; Dual use

Introduction

Synthetic biology is a techno-scientific discipline with the declared goal of rationally engineering biological systems utilizing approaches similar to those used to design bridges and send people to the moon. Inter alia, synthetic biologists aim to create redesigned or wholly novel organisms that serve human purposes. Proponents of synthetic biology predict it will yield products with valuable applications in medicine (e.g., new vaccines and other pharmaceuticals), energy (e.g., biofuels), environmental remediation (e.g., environmental cleansers), and agriculture (e.g., hardier crops) (Evans and Selgelid 2014). There are also hopes it will create new jobs and boost the global economy in the process (Presidential Commission for the Study of Bioethical Issues 2010).

Historical Origins: Recombinant DNA

Synthetic biology represents an extension of the life science revolution that began in the 1970s with the development of recombinant DNA (rDNA) – which, inter alia, involves insertion of isolated segments of DNA into the genomes of living cells (i.e., traditional "genetic engineering"). Early users of rDNA recognized the potential for rDNA methodologies to revolutionize the life sciences, through the introduction of rDNA derived from one organism into another (Berg et al. 1975).

The ability to combine the genetic information of radically different organisms, from early on in

the rDNA revolution, was recognized as a potential hazard to public safety. This potential motivated the Asilomar Conference on Recombinant DNA Molecules in 1974, where participants agreed that:

> most of the work on construction of recombinant DNA molecules should proceed provided that appropriate safeguards, principally biological and physical barriers adequate to contain newly created organisms, are employed (Berg et al. 1975).

Following this declaration, new scientific techniques, laboratory procedures, educational tools, and biosafety oversight measures were developed and implemented to reduce risks posed by rDNA experimentation.

The advent of rDNA paved the way to a host of new biotechnologies involving the transfer of genes between organisms. In 1978, for example, scientists inserted the gene that expresses human insulin into *E. coli* bacteria, producing a strain of *E. coli* that generated "synthetic" insulin (Presidential Commission for the Study of Bioethical Issues 2010).

Polish geneticist Waclaw Szybalski described rDNA, and its associated technologies, as a field of endeavor with "hardly any limitations to building 'new better control circuits' or. . .finally other 'synthetic' organisms, like a 'new better mouse,'" describing this new field as "synthetic biology" (Shatkai and Kohn 1974). This reference to "synthetic biology" is one of the earliest recorded uses of the term, though it differs in important ways from the use of the term today. Szybalski only mentions the modification of naturally occurring organisms, whereas contemporary synthetic biologists envision, among other things, design/creation of new ones.

Synthetic Genomics

Synthetic genomics, developed at the turn of the twenty-first century, combines advanced methods for the chemical synthesis of DNA sequences (i.e., building DNA sequences from chemical components) with computational techniques for their design, allowing scientists to construct genetic material that would be impossible or impractical using previous biotechnological approaches (such as rDNA) (Garfinkel et al. 2007). While the Human Genome Project was facilitated by, and itself accelerated, revolutionary developments in rapid DNA sequencing technology during the 1990s, synthetic genomics involves the more recent revolutionary development of technology that enables increasingly rapid synthesis of increasingly large DNA sequences.

It is now even possible to synthesize entire genomes of some viruses and bacteria. The synthesis of virus genomes, furthermore, in some cases enables artificial synthesis of actual "live" viruses. In 2002, for example, scientists funded by the US Army Defense Advanced Research Projects Agency (DARPA) used synthetic genomics to synthesize a polio virus. Following the published map of the polio (RNA) genome, which is published on the Internet, they purchased and strung together corresponding DNA sequences. The addition of the synthesized genome to "cell juice" (a solution containing cellular ingredients but no live cells) resulted in a "live" virus that paralyzed and killed mice (Selgelid and Weir 2010).

Advances in synthetic genomics have generated concern about the ethics of synthesizing naturally occurring dangerous pathogens (that are not otherwise easy to access) or modifying existing ones to increase their virulent properties. It is especially worrisome that aspiring bioterrorists might use synthetic genomics for such purposes. In 2005, for example, scientists used techniques of synthetic genomics to reconstruct the 1918 H1N1 (or "Spanish") influenza virus and published details about how they did so. Given that this virus killed an estimated 20–100 million people (in 1–2 years) (Crosby 1989), the paper in question was reviewed by the US National Science Advisory Board for Biosecurity (NSABB). Though NSABB approved publication, this decision has been subject to controversy – because the published study might provide a "recipe" or "blueprint" for a biological weapon of mass destruction.

Synthetic Biology and the New Synthetic Life Sciences

Where contemporary synthetic biology departs from its forebears is its approach to biology as a

form of engineering. Its focus is the application of engineering principles to biology in order to redesign existing biological parts, systems, and organisms or to design entirely new biological parts, systems, and organisms (Samuel et al. 2009). Synthetic biology is thus a true engineering discipline in contrast to traditional "genetic engineering" – which, despite its name, did/does not so explicitly apply engineering principles (or involve actual engineers). While synthetic biologists may use synthetic genomics to partially or wholly synthesize genetic material of redesigned or wholly novel life forms, a distinctive feature of synthetic biology (vis-á-vis synthetic genomics) is the aim to create life forms substantially different from those that already exist (or existed). Synthetic genomics, as illustrated in the examples above, could also be used merely to create already – or previously – existing life forms. Though closely related, synthetic genomics and synthetic biology should thus be distinguished (Samuel et al. 2009). Following Samuel et al. (2009), we use the term "synthetic life sciences" to refer collectively to both synthetic genomics and synthetic biology.

To advance their aims, contemporary synthetic biologists pursue a number of different projects. One of the most visible of these projects is the creation of libraries of standard biological "parts" and "devices" (where devices are composed of multiple parts) with known/predictable functions or properties. The Registry of Standard Biological Parts is one such catalogue that contains, to date, more than 3,400 parts. Much like LEGO pieces – or resistors, transistors, amplifiers, etc., in electronics – these so-called "biobricks" are meant to serve as the building blocks of synthetic biology.

Another high-profile activity is the International Genetically Engineered Machine (iGEM) Competition. The iGEM Competition is an annual competition, due to enter its thirteenth year in 2016, where teams compete to create the most innovative design based on a predetermined toolkit of biological parts drawn from the Registry (Carlson 2010). Though the iGEM Competition was initially limited to Massachusetts Institute of Technology students, it is now a cosmopolitan event drawing a range of competitors from high school age onward, producing a new base of users that – like young computer programmers – are introduced to the field of synthetic biology at a young age.

Synthetic biologists also seek to design microbial pathways to generate chemical compounds (or their precursors). These projects seek to render biological systems as "microbial chemical factories" with applications in energy, industry, and medicine. In what is arguably the most successful of these projects, scientists at the University of California, Berkley, created novel strains of *E. coli* (bacteria) and *Saccharomyces cerevisiae* (yeast) that both produce artemisinic acid, a precursor to the antimalarial compound artemisinin (which is naturally derived from the wormwood plant, via somewhat difficult/expensive processing).

Some synthetic biologists aim to create a "minimal genome": constructing the simplest possible genetic sequence to promote self-sustaining life. The resulting "minimal microbe" (i.e., the most basic/simple life form) would form the "chassis" into which engineered biological systems serving known/desired functions (such as chemical production) would be added to produce biological devices.

These advances, however, have been surrounded by controversy. Most recently, for example, scientists at the University of California, Berkeley and Concordia University in Montreal engineered a strain of yeast to produce the precursor to opioids found in the poppy plant (DeLoache et al. 2015). Though potentially an important new method for creating painkillers for people recovering from surgery or with chronic pain conditions, some are concerned that the new strain will provide an easy, scalable, and portable means to create illicit drugs like heroin, enabling drug cartels to function with increased ease (Oye et al. 2015).

Ethical Issues

Despite its potential benefits, synthetic biology raises numerous ethical concerns. The ethical

importance of synthetic biology is illustrated by the fact that the first report of Obama's Presidential Commission for the Study of Bioethical Issues focused on synthetic biology in particular (Presidential Commission for the Study of Bioethical Issues 2010).

Because synthetic biology is an emerging field, much of the debate, to date, is prospective and based on the potential benefits – or harms – of future developments. The ethics of synthetic biology thus largely turns on questions about how to prevent synthetic biology from causing harm and/or perpetuating injustices, without unduly impeding the progress of a field that has the potential to significantly benefit humanity.

Intellectual Property

Developments in synthetic biology raise questions about intellectual property rights: e.g., should new life forms created via synthetic biology be patented and/or patentable (see ▶ Patenting)? Debates about patents involve questions about:

- Whether developers should be able to patent life forms at all
- Whether patents, in the context of synthetic biology, foster or hinder scientific progress
- How patents might limit access to (e.g., pharmaceutical) products of synthetic biology by those who need them most

Patenting Life

There is debate about what life forms, or what aspects of biological devices, should or should not be patentable. Such controversy extends from older debates in other past and current life science areas such as rDNA, genetically modified organisms, and DNA sequencing. Organisms, some argue, cannot be patented because living things should not be "owned" in the relevant sense. It has been argued that owning an organism in the sense of a monopoly derived from a patent is different from ordinary ownership of dogs, cats, or cows – because a patent involves treating organisms as *mere* property for the purpose of profit, rather than as creatures with their own interests (e.g., (Hettinger 1994)).

The validity of such concerns, of course, depends on the organism, or what *about* the organism, a patent seeks to monopolize. A gene, for example, has no "interests" in the sense that a sentient creature might. Moreover, single-celled organisms may not have interests worthy of independent consideration, or to the same degree as animals. Most animals, in turn, may not have the same interests as, or interests warranting the same consideration as those of, humans. Proponents of the view that life cannot be patented, nonetheless, may counter that even simple creatures have some central interests worthy of independent consideration. Finally, the implications of owning not just *a* life, but claiming property rights over an entire *class* of life, are subject to ethical debate (Hettinger 1994).

These arguments are most famously articulated by religious authorities and predate synthetic biology. In the 1980s, the World Council of Churches, National Council of the Churches of Christ in the USA, Roman Catholic Church, and others against gene patenting released a series of reports decrying the patenting of genes. By 1995, almost 200 religion leaders from around the world endorsed a press conference by Jeremy Rifkin named the "Joint Appeal Against Human and Animal Patenting." Though many of these organizations emphasized that they did not hold an in principle stance against biotechnology, they were united in opposition to the patenting of living organisms (Hason 2001).

Patents and Scientific Progress

Biomedical innovation can be very expensive. In the case of pharmaceuticals, the cost of research and development is estimated to be between $161 million USD and $1.8 billion USD per drug (Morgan et al. 2011). Patents arguably allow developers to recoup the costs of investment, providing an incentive to participate in synthetic biological research and driving further scientific

progress. Patents, however, do not drive scientific progress in all cases. Costly licensing fees may limit participation in the biotechnology enterprise to only the most powerful (and wealthy) actors. Those who aim to participate may find their work subject to intellectual property (IP) litigation and/or face the logistical burden of navigating a landscape filled with competing IP claims by different firms. This may discourage researchers from pursuing profitable avenues of inquiry (that could be beneficial to humanity).

In synthetic biology, excessive patenting may result in a "patent thicket," in which innovation stalls in the face of multiple competing patent requirements for each new device. Such concerns have led some to advocate against patenting of organisms, arguing that a "biological commons" will best drive innovation (Carlson 2010). Making the life sciences open to all, according to this view, will encourage/enable a larger number of researchers to participate. Rather than having to create large incentives for actors, simply lowering access barriers to participation may promote pursuit of the field.

Positions for and against patents are not binary, however. Many industries exist and flourish with a hybrid of legal and social constructs that enable the enforcement of property rights, open communal endeavors, and other important aims – all at the same time. While there are extreme positions one can take between advocacy of patents and "open-source" solutions to the ownership of synthetic biological knowledge, it is also possible to accept a version of synthetic biology that incorporates both property rights and community interests – i.e., some products might be patented, while others are held in "biological commons." What version of synthetic biology we *ought* to accept, in the end, will partly turn on empirical questions about what are the best means to achieve legitimate social goals, including the goal of scientific progress.

Limits to Access

Appeals to scientific and technological progress, however, raise questions about who will, and/or should, benefit from science and technology. It is a well-known phenomenon that increased disparities between rich and poor have historically come hand in hand with scientific and technological advance (Farmer 2005). A concern with patents is that if the balance between property rights and societal benefit is struck too far in the direction of property rights, a situation may arise where the products of synthetic biology will primarily benefit those who are already well off. At present, only a minority of healthcare research spending is focused on afflictions responsible for the majority of the global burden of disease and vice versa. This is, in part, because current incentives for pharmaceutical innovation (i.e., via the patent regime) disproportionally favor the creation of products that meet the wants and needs of those who can pay – i.e., those who are relatively wealthy.

Neglected or "orphan diseases" in both developed and developing nations are chronically underfunded. An open, vibrant synthetic biology could enable the development of valuable drugs to combat diseases that kill the most people worldwide. By protecting synthetic biology against patents, it is argued that researchers who are interested in solving health problems of the poorest countries on earth – particularly those researchers within those countries – will be able to access technology and knowledge at little cost. Whether the promised benefit that synthetic biology will solve some of the world's most urgent public health problems is actually realized may thus largely depend on how debates about intellectual property rights are resolved.

Broader Justice Concerns

The ethics of synthetic biology as it pertains to neglected diseases, and groups historically made vulnerable by poverty, dovetails into broader concerns of justice implications of synthetic biology. For example, it has been argued that the synthetic production of artemisinin may put (disempowered) farmers growing wormwood out of business (Samuel et al. 2009). Whether or not

patented, this new form of production – providing a cheap alternative to an otherwise time- and labor-intensive process – may undercut traditional methods on which underrepresented or disenfranchised producers rely. This new technology may thus create new, or exacerbate existing/prior, injustices.

Such issues are not unique to synthetic biology. With increasing regularity, technological innovation undermines older methods of production. Mass production techniques undermined the skills of the artisan in areas as diverse as food, furniture, and metalworking. The personal computer industry has made it easier to access information cheaply, diminishing the need for libraries, or centralized news publishing. Robots replace laborers in numerous industries. The biological revolution may disrupt old methods of creating medicines, but it does not necessarily follow that this is a bad thing or that harms/injustice would outweigh potential benefits.

Creating Life

Some may be ethically worried about the artificial creation of life and/or new life forms – and about whether or not such an activity involves human hubris and/or "playing God" and/or is problematic because it is "unnatural." These kinds of concerns are commonly raised about numerous other developments in biotechnology. Concerns about hubris in the context of synthetic life sciences may ultimately turn on the biosafety and/or biosecurity dangers of synthetic life sciences (discussed below). With regard to the concern that synthetic life sciences involve "playing God," common responses are that this kind of objection will only appeal to those who believe in God and/or that it is not obvious that God would not want us to engage in this kind of activity (insofar as God, if He/She exists, has apparently given us the ability to do so). It is noteworthy that, according to the Presidential Commission for the Study of Bioethical Issues (2010), no established religion has officially expressed opposition to synthetic biology on such grounds.

Biosafety and Biosecurity

Newly created life forms might damage human health and/or ecosystems if they escape from labs or are intentionally released into the environment. A paradigm, if extreme, example is the fear of a "gray goo" scenario, where unpredictable new organisms reproduce out of control and consume the planet's resources (including humans or at least resources needed by humans). This kind of objection has also been raised about genetically modified organisms more generally. Defenders of synthetic biology have responded that such dangers could be avoided by designing synthetic organisms to contain "suicide genes" or by designing them to be dependent on artificial nutrients that would be unavailable if not intentionally provided by humans (Presidential Commission for the Study of Bioethical Issues 2010). Such protection measures, however, might not reliable in the context of reproducing/mutating/evolving organisms.

Even if the biosafety risks posed by synthetic biology are not so extremely catastrophic, there is still concern over the proliferation of risk from large numbers of small-scale "garage" labs. Part of what makes synthetic biology revolutionary is the development of more advanced technologies with which to conduct life science research at low cost. As biotechnology becomes a commercial venture, privately owned and run laboratories are emerging, conducting small-scale development of novel biotechnologies (Carlson 2010). While it is hoped (by some) that these "DIY" (do it yourself) or "garage biologists" will accelerate progress in synthetic biology, the possibility of many more laboratories (not subject to ordinary institutional oversight) implies a corresponding increase in the potential for serious laboratory accidents and laboratory-acquired infections. While not necessarily as catastrophic as "gray goo" scenarios, an increased potential for accidents involving novel biological agents could cause major environmental damage, threaten public health, and/or burden healthcare systems (Evans and Selgelid 2014).

With regard to biosecurity, the concern is that synthetic life sciences have "dual use" potential (see ▶ Dual Use). Though synthetic biology is

poised to benefit humanity, it could also be used by malevolent actors to cause grave harm. In particular, the techniques of synthetic biology might enable aspiring bioterrorists to design and create new highly contagious and deadly "designer pathogens" to be used as biological weapons, and/or that mere synthetic genomics could enable artificial creation of already existing pathogens (such as smallpox or Ebola) that bioterrorists might not otherwise be able to access (easily).

Given the potentially severe consequences that could result from malicious use of synthetic life sciences, it has been argued that increased oversight of research and/or publication of potentially dangerous discoveries may be necessary, that science codes of conduct for scientists (explicitly addressing dual use issues) should be adopted, and/or that scientists should be further educated about the dual use phenomenon and ethics. The degree of restrictive regulation that should be adopted in response to such dangers depends on the weight of the value of an open, unrestricted life sciences leading to beneficial progress, compared to the risks posed by intentional harms caused by products of synthetic biology. Some, meanwhile, have downplayed concerns about biosecurity by arguing that it is unlikely that humans will be able to create pathogens more dangerous than those that arise naturally.

The methods for weighing these values, and who ought to bear the burden for demonstrating the value of pursuing dual-use research, are subject to debate. In cases where technology poses a potential catastrophic risk – e.g., a bioweapons attack that harms millions – some appeal to the "precautionary principle," a name for a cluster of different strategies for approaching risk in situations involving uncertainty. As a *replacement* for a typical cost-benefit analysis, in which a particular course of action is assessed in terms of the probability and magnitude of the costs and benefits it incurs, strong versions of the precautionary principle would lead decision-makers to reject a course of action that incurs the possibility of some serious harm occurring. In the case of dual-use synthetic biology research, strong versions of the precautionary principle would arguably lead to some research (or publication thereof) being prohibited on the grounds that a catastrophic bioterror attack could result (Clarke 2013).

A strong precautionary principle applied *ex ante* to the recent synthesis of opioids, for example, may entail restricting access to the experimental results, if not prohibiting future experiments, until the potential risks (e.g., use by cartels) could significantly reduced or eliminated. This (very) strong precautionary principle would not entail balancing the risks of this synthetic biological technology against any potential benefits, but rather only protecting against future harms.

Opponents of strong versions of the precautionary principle typically claim that it is untenable because almost every option will carry some serious risk. The aim to avoid all serious risks can thus lead to a form of choice paralysis. This undermines (strong versions of) the precautionary principle's ability to guide decision-making. Alternatives given by proponents of the precautionary principle include setting a threshold for what counts as a "serious" risk, to reduce the chance of paralysis, or adopting a weaker version of the precautionary principle that instead focuses on who ought to bear burden of proof for ensuring that certain serious harms are prevented or mitigated.

Supporters of a weaker precautionary principle, applied again to the synthesis of opioids, might call for a temporary moratorium on research – or on particular kinds of research – until appropriate risk mitigation processes can be implemented. Four recommendations given by Oye et al. (2015) for the management of the risks of synthetic opioids are (1) engineering yeast strains to make them less appealing to criminals, (2) screening commercial DNA sequences, (3) enhancing physical security around sites using modified strains of yeast, (4) and extending existing law to criminalize the unauthorized distribution of opioids. They argue that the research seeking to fulfill the first of these recommendations might be permitted – but other kinds of research should be delayed until all four recommendations are satisfied.

Alternatives often try to find a happy medium: the chair of the Presidential Commission, Amy Gutmann, favors a principle of "responsible stewardship" in synthetic biology. The Commission recommended continual monitoring of synthetic biology, to rapidly assess new risks emerging from the field. Ideally, this monitoring will be followed with strategies to mitigate risks when they present a serious threat to human health. Rather than being a single, *ex ante* risk assessment of synthetic biology, responsible stewardship takes the form of a continual assessment (Presidential Commission for the Study of Bioethical Issues 2010).

Still others might prefer a threshold account, whereby one switches from classical to precautionary approaches to risk when a probable outcome becomes sufficiently harmful. Under such an approach, two otherwise similar pieces of research or technology would be treated differently if one had the potential to be used to cause mass harm, even if the likelihood of mass harm was low enough that the research was still expected to cause more benefit than harm in the long run. Synthetic opioids might not have such potential for mass harm, but research that could create botulinum toxin in an easily weaponizable form might be treated differently even if, on analysis, it was expected to have long-term net benefits. This is because botulinum toxin, as a potential biological and toxin weapon, could be used to harm hundreds of thousands of people in a single well-executed attack. Under a threshold account, this additional, high-magnitude harm might be sufficient to initiate a switch from a mere weighing of benefits against risks to a precautionary approach that seeks to protect against a particularly large harm before research moves forward.

Conclusion

Synthetic biology has the potential to produce important developments in medicine, agriculture, and energy production. How we should reconcile the promise of these advances with the perils that could arise from the discipline is the subject of continued debates concerning intellectual property rights, the importance (and best ways to achieve) scientific progress, justice, creation of life, biosafety, and biosecurity. These debates ultimately turn on questions regarding how we define the benefits of this emerging field, how best to achieve those benefits, what trade-offs are acceptable in this pursuit, and whose benefit should carry moral weight in decision-making pertaining to synthetic biology.

Cross-References

- ▶ Dual Use
- ▶ Genetic Modification (GMOs): Animals
- ▶ Indigenous Knowledge
- ▶ Patenting
- ▶ Precautionary Principle

References

Berg, P., Baltimore, D., Brenner, S., Roblin, R. O., & Singer, M. F. (1975). Summary statement of the asilomar conference on recombinant DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America, 72*(6), 1981–1984. doi:10.2307/64627?ref=no-x-route:608f8cefda0161234b77e36dd8a58c58.

Carlson, R. H. (2010a). *Biology is technology: The promise, peril, and new business of engineering life*. Cambridge, MA: Harvard University Press.

Clarke, S. (2013). The precautionary principle and the dual-use dilemma. In M. J. Selgelid & B. Rappert (Eds.), *On the dual uses of science and ethics* (pp. 223–233). Canberra: ANU E Press.

Crosby, A.W. (1989). *America's forgotten epidemic: The influenza of 1918*. Cambridge, UK: Cambridge University Press.

DeLoache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J. J., & Dueber, J. E. (2015). An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nature Chemical Biology*. doi:10.1038/nchembio.1816.

Evans, N. G., & Selgelid, M. J. (2014). Biosecurity and open-source biology: The promise and peril of distributed synthetic biological technologies. *Science and Engineering Ethics, 41*(4), 1065–1083.

Farmer, P. (2005). *Pathologies of power*. Berkeley: Univ of California Press.

Garfinkel, M. S., Endy, D., Epstein, G. L., & Friedman, R. M. (2007a). Synthetic genomics: Options for governance. *Industrial Biotechnology, 3*(4), 333–365. Retrieved from http://www.jcvi.org/cms/research/projects/syngen-options/publications/.

Hason, M. J. (2001). Religious voices in biotechnology. In M. J. Hason (Ed.), *Claiming power over life* (pp. 72–111). Washington, DC: Georgetown University Press.

Hettinger, N. (1994). Patenting life: Biotechnology, intellectual property, and environmental ethics. *The Boston College Environmental Affairs Law Review, 22*, 267.

Morgan, S., Grootendorst, P., Lexchin, J., Cunningham, C., & Greyson, D. (2011). The cost of drug development: A systematic review. *Health Policy, 100*(1), 4–17.

Oye, K. A., Lawson, J. C. H., & Bubela, T. (2015). Drugs: Regulate "home-brew" opiates. *Nature, 521*(7552), 281–283. doi:10.1038/521281a.

Presidential Commission for the Study of Bioethical Issues. (2010). New directions. Washington, DC.

Samuel, G., Selgelid, M. J., & Kerridge, I. (2009). Managing the unimaginable. Regulatory responses to the challenges posed by synthetic biology and synthetic genomics. *EMBO Reports, 10*(1), 7–12.

Selgelid, M. J., & Weir, L. (2010). Reflections on the synthetic production of poliovirus. *Bulletin of the Atomic Scientists, 66*(3), 1–9.

Shatkai, A., & Kohn, A. (1974). Control of gene expression (p. 434). Presented at the Oholo Biological Conference on Strategies for the Control of Gene Expression, Zikhron Yaaqov, Israel.

Further Readings

Carlson, R. H. (2010b). *Biology is technology: The promise, peril, and new business of engineering life*. Cambridge, MA: Harvard University Press.

Garfinkel, M. S., Endy, D., Epstein, G. L., & Friedman, R. M. (2007b). Synthetic genomics: Options for governance. *Industrial Biotechnology, 3*(4), 333–365.

Presidential Commission for the Study of Bioethical Issues. (2010). New directions. Washington, DC.

Tucker, J. B., & Zilinskas, R. A. (2006). The promise and perils of synthetic biology. *New Atlantis, 12*(1), 25–45.