

Ethics in e-trust and e-trustworthiness: the case of direct computer-patient interfaces

Philip J. Nickel

Published online: 7 April 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract In this paper, I examine the ethics of *e-trust* and *e-trustworthiness* in the context of health care, looking at direct computer-patient interfaces (DCPIs), information systems that provide medical information, diagnosis, advice, consenting and/or treatment directly to patients without clinicians as intermediaries. Designers, manufacturers and deployers of such systems have an ethical obligation to provide evidence of their trustworthiness to users. My argument for this claim is based on *evidentialism* about trust and trustworthiness: the idea that trust should be based on sound evidence of trustworthiness. Evidence of trustworthiness is a broader notion than one might suppose, including not just information about the risks and performance of the system, but also interactional and context-based information. I suggest some sources of evidence in this broader sense that make it plausible that designers, manufacturers and deployers of DCPIs can provide evidence to users that is cognitively simple, easy to communicate, yet rationally connected with actual trustworthiness.

Keywords Direct computer-patient interfaces · e-trust · e-trustworthiness · e-health · Ethics of biomedical engineering · Evidentialism · Health care ethics · Obligations of designers · Trust · Trustworthiness

Introduction

In this paper I use notions of *e-trust* and *e-trustworthiness* to make an ethical argument about the design of information and communication technology (ICT) in health care. As I define it, *trust* is an attitude of willingness to rely on another person or entity to perform actions that benefit or protect oneself or one's interests in a given sphere of activity, together with a *normative expectation*: the person or entity *should* perform in a particular way. In *e-trust* the thing trusted is an ICT system consisting of computers, networks and operators. *Trustworthiness* is the counterpart of trust, a characteristic of a trusted person or entity such that it is likely to perform as expected and that it meets the normative expectations of trust. *E-trustworthiness* is this characteristic as applied to ICT systems. In what follows, I assume that technological artifacts and systems can be proper objects of trust. I explain and defend this view elsewhere (Nickel 2012).¹ I will freely use the terms trust and trustworthiness to refer to e-trust and e-trustworthiness in what follows.

I look particularly at direct computer-patient interfaces (*DCPIs* for short), computer systems which diagnose, advise and even treat patients directly by means of ICT. Direct computer-patient interfaces collect patient data, draw inferences from that data, and deliver information back to the patient on the basis of these inferences, assisting or replacing the function that a physician usually performs. For example, online health websites that take information from patients and deliver diagnoses or

P. J. Nickel (✉)

Department of Philosophy and Ethics, School of Innovation Sciences, Eindhoven University of Technology, Postbus 513, Den Dolech 2, 5600 MB Eindhoven, The Netherlands
e-mail: p.j.nickel@tue.nl

¹ Taddeo (2009) goes further and allows for the possibility that electronic artificial agents can actually have the attitude of e-trust, as well as being e-trustworthy.

recommendations for physician consultation are a form of DCPIs.² Direct computer-patient interfaces can also operate within and under the supervision of a clinical facility. More complex DCPIs have other features such as linkage with imaging or diagnostic equipment or biological test results, expert (human) review of the results, artificial intelligence (e.g., revision of inferential algorithms in light of new data), integration with medical records systems, etc. Direct computer-patient interfaces differ from telemedicine, in which ICT is used as a *medium* for medical care, in that DCPIs take over some of the intellectual tasks of the physician. Direct computer-patient interfaces are also more than mere health information sources, because they gather information about patients and respond on the basis of that information.

In what follows, I argue that the designers, makers and deployers of DCPIs have an ethical obligation to provide sound evidence to patients of these systems' trustworthiness. There are some reasons for initial skepticism about this claim. First, it is natural to assume that the main obligation a designer or manufacturer has regarding the trustworthiness of its products is to make the products themselves reliable, since a reliably functioning product is the defining goal of designers' and manufacturers' activity. In a discussion of hip replacement system design, for example, John Fielder argues that engineers' primary obligation is to make safe, well-functioning products. He also argues for a secondary, positive communication-oriented obligation: in case there are any known defects, use restrictions, or unsafe aspects of the product, the engineer is required to disclose these fully and promptly (Fielder 1992, discussed in Vallero 2007). But no mention is made of the need to communicate additional sound trust-related information to physicians or patients *in cases where the products are reliable or where it is not known that they are unreliable*. Similarly, a recent discussion of implantable heart transplants focuses on establishing standards for device reliability and disclosure of faults in medical devices, rather than on providing evidence for trust (Myerburg et al. 2006). One might reasonably suppose that DCPIs are similar to these other health devices: design of a reliable product and disclosure of any known faults or defects demarcate the limit of manufacturer obligations.

Moreover, positive ethical obligations are usually more difficult to justify than negative ones. (By 'positive ethical obligations', I refer to obligations most naturally expressed by describing what must be done, rather than what is forbidden and must be avoided.) Judy Thomson (1971), for example, famously argues that the positive obligation of laypersons to provide life-saving assistance to others does

not hold in cases where it is inconvenient to do so. Positive obligations take up valuable resources. One should be wary of introducing new obligations that impose burdens of action on already busy, morally engaged people.

Despite these considerations, I argue that some kinds of products must also be accompanied by genuine evidence of their trustworthiness. Direct computer-patient interfaces are an especially interesting case for three main reasons. First, there is widespread consensus concerning the ethical values that apply to medical professionals in their treatment of patients. It is widely agreed that medical professionals have a positive obligation to respect the autonomy of their patients (the Principle of Autonomy), to obtain patients' informed consent for medical procedures (Informed Consent), and always to act for the sake of their patients' benefit (Beneficence) (Beauchamp and Childress 2009). What would be optional, supererogatory or only "imperfectly" obligatory actions for ordinary people, such as making substantial time sacrifices for the sake of another person's health, are strictly or perfectly morally obligatory for doctors and considered part of their role obligations. Although the interpretation of Autonomy and Beneficence is sometimes disputed and conflicts between them sometimes difficult to resolve, the principles themselves are widely accepted and standardly taught in medical school curricula. Consensus is also emerging that trust is a central value of the clinician-patient relationship (*ibid.*, pp. 40–41).

Second, ICT permeates the practice of health care, making it unclear how the standard medical ethics framework ought to be adapted to situations in which computers mediate or replace relationships between clinicians and patients. People seeking information about health care, including sick people, often turn to the Internet for information (Uden-Kraan et al. 2009). Health care facilities test systems that allow patients to respond to Internet-based questionnaires about their health and receive tailored therapeutic feedback (Mangunkusumo et al. 2005). Intelligent computers can also interact with patients to obtain informed consent to therapies (Dunn et al. 2001; Anonymous 2009) and in principle even administer these therapies (Selmi et al. 1991; Bobylev et al. 1997). In the future, it is not unrealistic to suppose that many or all of these functions could be integrated into a single interface. How should we adapt medical ethics and the conception of trustworthiness to these developments?

Third, the use of DCPIs in particular brings the possibility of significant benefits and risks. On the one hand, the resource of clinicians' time is a crucial bottleneck in the availability of health care. Computers can make health care more widely available and cheaper, spending more time on communication with patients and doing so in more comfortable and convenient times and places. But on the other hand, there can also be serious risks associated with DCPIs.

² An example is EasyDiagnosis at <http://easydiagnosis.com> Accessed 26 January 2011.

This can be seen by drawing a comparison with the risk scenarios that experts have identified in telemedicine, a technology with many functional similarities to DCPIs (Stanberry 2001; Duplaga and Zielinski 2006). Patients entrust confidential information to these systems, such as information about symptoms and identifying personal information. Confidential diagnostic data can also be generated by these systems. Moreover, patients may base crucial medical and non-medical decisions on the diagnosis and advice they receive from such systems. Furthermore, patients may sometimes perceive these systems as replacements for traditional medical consultation and avoid seeking further medical help.

Although others have brought the notion of trust to bear on ethical issues in health care (O'Neill 2002; Illingworth 2005), these attempts can be improved upon in two philosophical respects. First, the account of what trust is can be sharpened so as to serve as a plausible shared starting point for ethical arguments, making it easier to identify where trust can be found or how it can be stimulated or discouraged.³ Second, it can be made clearer why and under what circumstances trust is epistemically and ethically justified and when it is demanded or made salient by the circumstances.

Evidentialism about trust

I frame my argument with an elementary conceptual distinction. Theorists of trust such as Russell Hardin (2006) have noted that trustworthiness and trust are different (though related) concepts. Trustworthiness is the property of a person (or in the broader sense used here, an artifact or system) such that its performance can be relied upon, and such that it meets the normative expectations of potential trustors. Trust, on the other hand, is an attitude taken by people toward that entity, of willingness to rely on it. Whereas trustworthiness is possessed by the object of trust, trust is possessed by the person who trusts.

With this distinction in place I can now state the first of three propositions to be defended in what follows, that trust should be based primarily on evidence of trustworthiness (what I call *evidentialism* about trust). Evidentialism is a view in the traditional philosophical debate about the “ethics of belief” holding that one’s belief states should conform to the available evidence.⁴ Evidentialism about trust holds that trust should be based on evidence that the

trusted entity will perform as anticipated and meet the trustor’s normative expectations. It contrasts with pragmatism, the view that other kinds of reasons (such as considerations of desirable consequences) are appropriate basic reasons for trust. In the realm of health care, a pragmatist might hold that patients should trust whenever doing so is good for their health or for the optimal (fair, profitable, etc.) functioning of the health care system, or perhaps whenever they have no other good option. These claims conflict with evidentialism.

I offer two main philosophical arguments in favor of evidentialism about trust. The first is an adaptation of a familiar style of argument called the “wrong kind of reason” argument. As Pamela Hieronymi (2005) explains the point, reasons are *considerations that bear on a question*. Different kinds of questions require different kinds of considerations to answer them. For example, we must distinguish between the question “Is it true that P?” and “Would it be good to have the belief that P?” (for example, “Is it true that I will survive the surgery?” vs. “Would it be good to believe that I will survive the surgery?”) Some kinds of considerations that bear on the second question do not bear on the first question: e.g., “If I believe that I will survive the surgery I will make my family happy” or “It is painful to think about the surgery.” If these are reasons at all, they are reasons for or against having a certain mental state, not reasons that bear on the probability that I will survive the surgery. Hence they provide the wrong kind of reason for answering the question to which that mental state intrinsically responds. Hieronymi holds that the attitude of trust is formed directly in response to the question, “Is S trustworthy?” Considerations about whether it would be good to have the attitude of trust do not all directly bear on this question (Hieronymi 2008). For example, if trusting a computer program would please the programmers, that might be a pragmatic reason to have the attitude of trust, but it would not make the program any more trustworthy and therefore would be the wrong kind of reason for trust.

The second argument is based on a minimal rationalist principle of morality, the Recognition Requirement (Nickel 2001). This requirement states that a decision is morally good only to the extent that one decides from a recognition of relevant reasons. Take an example of a medical decision made on somebody else’s behalf: suppose Betty is unconscious and Al, her designated proxy, must decide whether to elect a particular surgery as a treatment for her. Al should decide based on reasons such as Betty’s preferences and the risks and benefits of the surgery. Suppose

³ For example, Beauchamp and Childress’s discussion is based on a definition of trust as “a confident belief in and reliance on the moral character and competence of another person ...” and according to which trustworthiness is a moral virtue (*op. cit.*, p. 41). This contrasts markedly with social scientific conceptions of trust, which often hold that trust and trustworthiness are not inherently moral qualities.

⁴ Evidentialism about the ethics of belief is distinct from the view, also sometimes called evidentialism, that justification can be analyzed in terms of the evidence available to the subject. See Conee and Feldman (2004).

AI makes the correct decision, but not based on the right reasons: he flips a coin without even considering the reasons for and against the surgery. In that case, his action would not ordinarily be morally good. The important thing to notice here is that the Recognition Requirement has an informational aspect: in order for AI to recognize the right reasons, he needs to have relevant information available. If AI cannot find out what the surgery is or what Betty's preferences are, then he cannot make the decision on the right basis and therefore cannot act morally well, even if by accident he gets it right. AI can be excused for his poorly-grounded decision in the event that relevant information is totally unavailable, but the very fact that he must be excused proves that something has gone wrong. So in morally important decisions, having access to the right kind of reasons is a precondition for acting morally well.

The Recognition Requirement thus gives AI strong reason to be attentive to available information and seek out relevant information that is lacking. But as I argue in the next section, the Requirement does not only carry implications for AI. It also places an requirement on those on whom AI might rely for Betty's treatment. These people are in a position to provide some of the information AI needs in order to make a well-justified decision. In a responsible practice of health care, professionals (whether clinicians or engineers) will do what they can to provide AI with information that helps him meet his own informational burden.

Relevant evidence about the trustworthiness of a DCPI system for aspects of one's health care consists of information about its capacity to make true and accurate statements, to protect one's private data, to make appropriate diagnoses, etc. One way of presenting this information is to state the system's track record of success in similar cases or extensive clinical testing, in which the risks of inaccurate diagnosis, breaches of confidentiality and so on are estimated and this information presented to the patient. In medical contexts this is the normal way of implementing the Principle of Informed Consent. When patients are presented with the option of surgery or other medical treatment, they are given statistical information about the likelihood and severity of various risks associated with the treatment. This ensures that their decision is not arbitrary and that they take responsibility for its moral consequences.

There is an important worry here, however, that these normative requirements are unrealistic. Some psychologists studying human decision making have concluded that people make inferences on the basis of inaccurate and inconclusive evidence and that they are incapable of the rationality implied by evidentialism (Tversky and Fox 1995; Tversky and Kahneman 1992). There are also more specific doubts about the ability of people to make rational inferences about the impact of good and bad events on their

well-being (Wilson and Gilbert 2003), particularly in the health care domain (Ditto et al. 2005). This leads to the worry that it is impossible or impractical to meet the burden of evidentialism.

I have three main responses to this objection. First, although people may be decision-theoretically irrational in some contexts, one prevalent view in psychology is that they nonetheless have a "bounded rationality" that applies to certain contexts, enabling them to make quick, rational decisions much of the time (Gigerenzer and Selten 2002). Second, people with a better than average ability to make evidence-based practical decisions about their healthcare deserve to have good information on which to base their trust. And finally, the concept of evidence or rationality presupposed by those who question the ability of humans to decide rationally is too narrow. Evidence need not be conceived of so narrowly as only to include probabilistic information about risk. It can also include social, interactional, and contextual information. I describe some of these sources of information in the concluding section.

The obligation to provide evidence of trustworthiness

My second main proposition is that if the designer, manufacturer or deployer of a DCPI elicits patient trust concerning serious health matters, evidence of the system's trustworthiness should be made available. Weaker trust-oriented duties have been suggested by moral philosophers in the past. Tim Scanlon advocates a principle forbidding the intentional creation of false expectations in a subject: "One must exercise due care not to lead others to form reasonable but false expectations about what one will do when one has good reason to believe that they would suffer significant loss as a consequence" (1998, 300). But Scanlon does not suggest that the trusted party has an obligation to provide information about its trustworthiness to the subject. Why should there be such an obligation? As hinted in the previous section, my argument is as follows. If a morally good action requires access to relevant reasons for that action and if it is the case that one is well-positioned to make such reasons available to the subject, then it seems one has the opportunity to make morally good action possible. If one then fails to do so when given the opportunity, other things equal, it seems one has not acted rightly. Thus it appears that makers and deployers of DCPIs have ethical reason, other things equal, to make evidence about trustworthiness available to potential users of these systems.

In other cases where a product is offered to consumers, it is not generally assumed that they also have a *duty* (rather than just prudential reasons) to provide information about the trustworthiness of the product. This has a great deal to do with resource constraints and cognitive

limitations in information exchange. In practice the ability of people to make a well-informed judgment about trustworthiness is limited, yet they often must trust anyway (in the sense that for practical reasons they must rely on others and they accept this fact voluntarily). Even for people with multiple options, good information, and substantial resources for investigation, it may not make sense to spend much time reflecting on their trust because doing so is time-consuming and competes with other resources and priorities. Sick people in particular are beset with many other goals and demands and are often under particular strain because of illness. Although it may be rational for them to spend time investigating and weighing the reliability of their health care, they may have many other priorities.⁵ Furthermore, their ability to process this information may also be weakened by illness, fatigue or distraction. Health problems affect people without much regard to levels of education and informational access, so many people have a hard time understanding health information unless it is carefully communicated. In addition, clinicians have little time to communicate with patients (Østbye et al. 2005). Hence patients are put in a position to trust clinicians without having much information about trustworthiness.

Yet there is reason to think that providers of health care information and services have a special obligation to provide information. A distinctive fact about the ethics of health care is that health care providers have various *positive* duties to patients. It is commonly assumed that health care providers cannot meet their ethical obligations merely by “doing no harm” (National Commission 1979). They have a positive obligation of beneficence which motivates strict duties for health care providers to give positive assistance. While the makers of fruit juicers or paint primers might have only a weak obligation to provide evidence of the trustworthiness of their products, the makers of a DCPI technology are governed by more stringent background obligations to provide positive informational assistance.

Furthermore, if evidence can be made easier to understand, so that it takes less patient time to evaluate and becomes more widely accessible to those with cognitive limitations, then some of the obstacles that might otherwise stand in the way of meeting the obligation to provide evidence of the trustworthiness of DCPIs will be reduced. To some extent, DCPIs can increase the basis for trust by giving patients more information about trustworthiness. Because their use is not restricted by the limited resource of

physician contact time, they ease the problem of meeting the evidentialist requirement.

There are still substantial worries about the epistemic practicability of this obligation, however. Insofar as the DCPI itself is used to deliver information about its trustworthiness, there will be a circularity problem: the validity of the information about trustworthiness will only provide the right kind of reason if the DCPI is already trustworthy. Take an obvious example: if a DCPI has a “trust-page” on which the patient reads that the system was developed by the most knowledgeable experts in the field, this information can only be as reliable as the system itself. It seems that the patient will need *independent* information about the reliability of the system in order to have reason to trust it. There is also the problem of *discerning* valid health information sources from bogus sources of health information, designed by those with a purposeful intention to deceive, or with the sale of bogus medical services as their aim (“informational snake-oil”). Systems can fake the validation of outside health authorities, illicitly linking to legitimate websites or bypassing browser verification systems to create deceptive mirror certification sites. Some of the sources of evidence I mention in the next section can help address these problems.

What is good evidence of the trustworthiness of DCPIs?

In this final section, I discuss the nature of evidence of trustworthiness as provided by DCPIs, offering some preliminary ideas about how the obligation to provide such evidence can be met. My remarks are intended to show that fulfilment of the obligation is feasible, rather than to provide a detailed set of recommendations.

Evidence of trustworthiness consists of some available sign or phenomenon that makes it more likely that a desired performance is worth counting on and may be normatively expected. Consider again the example of a “trust page” on a DCPI website. Does the presence of this page make it more likely that the DCPI will produce truthful information about the patient’s condition and keep her personal information confidential?⁶ The relevant general question is: *does the total evidence presented to the user of a DCPI make it reasonable to expect it to perform and make its performance sufficiently likely that it is worth staking one’s actions on it relative to other salient options?* In what follows I will focus on those aspects of the total evidence

⁵ Evidence suggests that patients need time to absorb difficult and emotionally fraught information, and that they go through phases in which they do not want additional information (Leydon et al. 2000; Case et al. 2005).

⁶ Some signs are bad evidence even though they have a salient psychological impact. An early study of trust in online advice concludes that users trust websites with a good “look and feel,” that are “easy to navigate and free of errors and clutter” (Briggs et al. 2002, 330). But this leaves open the question whether an easily navigated website is more likely to be trustworthy than one that isn’t.

which are *interactional* (having to do with one's interaction with the system and its designers and especially its deployers) and on those which derive from the *socio-technical context* of deployment of the DCPI. These are the most helpful evidence sources for users of DCPIs and are also most likely to overcome the epistemic problems mentioned at the end of the previous section.⁷

Russell Hardin's notion of trust as "encapsulated interest" is a good place to start in the search for interactional evidence about the trustworthiness of DCPIs. According to Hardin (who is analyzing *interpersonal* trust rather than *e-trust*), the rationality of one's trust in another person depends on whether one's own interests are encapsulated in that person's interests. There are various reasons for interest-encapsulation, such as the desire for future interaction and exchange (reciprocal dependence), the concern of those involved for their general reputation (reputational staking), or the fact that one will be harmed or legally punished if one does not fulfill one's expectations (sanction threat) (Hardin 2006).

Reciprocal dependence is likely to be an *insignificant* source of evidence about the trustworthiness of DCPIs. The designers, manufacturers, and deployers of a DCPI are unlikely to be dependent on the future action of any one particular patient.⁸ The dependence relationship is unilateral, not reciprocal in the way needed to provide additional evidence of trustworthiness.

Reputational staking, on the other hand, is prevalent among e-health systems and can take a number of forms. If a DCPI is located in a clinical setting, the clinical institution itself stakes its reputation on the system. In web-based e-health applications, certification often consists of a label displayed when a user accesses a website, an "about us" page, or an institutional embedding of the site (Eysenbach 2000; Anonymous 2011). The visible presence of a link to a certifying governmental agency or independent professional organization makes it more likely that the DCPI is trustworthy because it ensures that the reputation of a recognized, independent institution is staked on the reliability and security of the software. For example, a telephone-based test for

depression is currently being offered in the Netherlands whose website bears the logo of the VU Medical Center, a major academic medical center.⁹ Through this visible sign, the VU Medical Center incurs some responsibility for the reliability of the test for depression. This sort of reputational staking is widely regarded to make a difference to trustworthiness (Coleman 1990; Pettit 1995).

There are significant skeptical worries about the ability to take advantage of such evidence, however. First of all, sometimes a whole sector is worthy of suspicion. For example, during the recent financial crisis certain highly dubious financial products were widely traded and regarded as legitimate by major banks and many in academia. Reputational staking by legitimate businesses and academic institutions did not guarantee trustworthiness. This seems to show that even certain sound financial products could not be trusted by non-experts because they could not reliably be distinguished from unsound products. Since there are comparable worries about the complicity of academia and regulators in the certification of pharmaceuticals and medical procedures, one might be concerned that no amount of reputational staking can sufficiently support a judgment of trustworthiness in DCPIs.¹⁰ Secondly, it is difficult for many people to distinguish between legitimate and illegitimate third party certifiers. And thirdly, any certification of a DCPI will be general and will not directly transfer to each configuration and application of the system because the specific configuration and application are sometimes questionable as well.

The first worry, concerning situations in which a whole sector of activity is untrustworthy, is the most difficult of the three to address. It is not always possible to provide valid evidence of the trustworthiness of reliable products. In cases where a whole sector is suspect, fundamental measures to reduce risk and provide better governance may be needed to reestablish and demonstrate trustworthiness, distinguishing good from bad products. Direct computer-patient interfaces and the e-health sector in general appear not to be in this dire state, but the decentralized nature of regulation in some parts of the sector may make it vulnerable.¹¹ It is widely agreed that trust is difficult to reestablish once it has been widely undermined (Walker 2006).

The second worry is that it is difficult for people to distinguish legitimate from illegitimate certifiers. As a first response, it is worth going back to the point that

⁷ It is important to note that the types of evidence discussed in this section may be insufficient in situations where there is a stringent burden to provide even more probing trust-related information to the users of DCPIs. For example, when using experimental DCPI systems to administer medical care, it may be necessary to explain the system, its clinical or therapeutic function, and any possible risks of its use in greater detail. (I wish to thank an anonymous reviewer for this point.) Indeed, in such cases it is probably not yet reasonable to have normative expectations about how the system should function, and so the system itself cannot be trustworthy, even if *per accidens* it is reliable and the experimental team is trustworthy.

⁸ For this very reason, Hardin himself confines the concept of trust to individual relationships. Most other social theorists take a broader view, however. See Möllering (2006) for an extensive discussion.

⁹ <http://www.nationaledepressietest.nl/>, Accessed 26 January 2011.

¹⁰ My thanks to an anonymous reviewer for posing these questions.

¹¹ Eysenbach (2000) emphasizes self-regulation and consumer education because he thinks we should accept this decentralization. With today's corporate consolidation of the web and greater concern for risk control, one might instead emphasize governance measures such as institutional oversight and embedding of services within established clinical institutions.

reputational staking can be combined with other means of demonstrating trustworthiness. The question is whether the *total evidence* presented to the user of a DCPI makes its trustworthiness likely. Threat of sanctions and contextual evidence can be combined with reputational staking, contributing to a pool of total evidence that supports trustworthiness. Governance structures such as laws, regulatory agencies and professional boards can be put in place to ensure that defective or fraudulent systems are detected and moral, institutional or legal sanctions are applied. The user, if made aware of this, can reasonably infer that the DCPI will comply with relevant standards. In addition, it is theoretically possible to certify the certifiers, iterating the same strategy of verification at a higher level. For evidential value it is best to make this as simple and concrete as possible, for example by making it possible to call a support person or an oversight authority with questions and concerns. And finally, it is possible to place the DCPI physically within a clinical context (e.g., a major public hospital) that is clearly legitimate.

The third concern is that evidence provided by certification and sanctions cannot “trickle down” to each specific configuration and use of a DCPI. For example, users might not know whether a general diagnostic system for depression can be used in their particular case (e.g., with a teenager). General interactional evidence of trustworthiness does not establish its trustworthiness for *this* use. This worry can be addressed both theoretically and practically. First, theoretically, although evidence of trustworthiness will always give some false positives (systems that are untrustworthy despite evidence to the contrary), this does not by itself undermine the value of evidence in establishing trustworthiness. The threshold for adequate evidence of trustworthiness is not so high that it must rule out every false positive; it need only ensure that the likelihood of unreliability is small. Users’ skeptical worries, in order to make additional evidence necessary, must be reasonable or well-grounded. Jonathan Adler, a well-known defender of evidentialism in epistemology, has argued that it is not, for example, the duty of a waiter to investigate every possible skeptical worry before asserting that a cup of coffee is decaffeinated, even if there are a few customers who have arrhythmia and could suffer as a result of being served caffeinated coffee (Adler 2002). To challenge the waiter’s assertion, the customer must have a valid reason for questioning this claim, such as that they themselves are arrhythmic or that they believe they saw the pots of coffee switched. The importance of Adler’s point here is that DCPIs should be actively responsive to *reasonable* concerns about the configuration and use of a system, but they need not respond to every conceivable worry about the system. This makes the burden feasible to bear. Practically, what this means is that the deployers of such systems must

provide a meaningful mechanism for tracking the success of the system in different implementations and for registering, evaluating and responding to the trust-related concerns of individual users and user groups. They should also be aware of the threat of illegitimate web-based DCPIs that may undermine the credibility of their systems, and they should take specific measures to differentiate their site from these other systems.

One of these measures, briefly referred to above, is to exploit a different type of evidence about the trustworthiness of DCPI systems: information derived from the socio-technical context of the system. As Carusi (2009) explains, discussing another kind of socio-technical medical system (technology that helps radiologists distribute and compare their interpretations of mammograms), the particular way an ICT system is contextualized and used is crucial to establishing trust among the system’s users. For example, a system which allows double-blind ICT-enabled confirmations of a mammogram result produces different feelings of trust than conventional double-readings of mammograms. Carusi points out that such contextual information is usually implicit as a reason for trust. Such implicit information, consisting of background beliefs or perceptions, can also serve as evidence (Adler 1990).

The socio-technical context makes a crucial difference to the additional evidence needed to establish trust in DCPIs as well. As mentioned above, a DCPI interaction under the supervision of a clinician in a health care facility inherits a great deal of evidential support from its socio-technical embedding. Strong evidence of trustworthiness is provided to the user by her warranted background beliefs about the reliability and professionalism of the institution. The user need not consider this evidence consciously in order to be more strongly warranted in relying on the system. In addition, there is little chance that a deceptive system has been smuggled into the facility, in the way that a bogus web-based DCPI might be confused with a legitimate one. Therefore, fewer skeptical concerns need to be addressed in order to provide adequate evidence for user trust.

It also follows that, *if* this context is taken as given, then it is much easier for the designers, manufacturers and deployers of the system to meet their evidentiary obligations. Indeed, they may need to do very little further to provide evidence of trustworthiness.¹² However, as e-health and other clinical innovations such as home health care robots change the boundaries of clinical care (Coeckelbergh 2010), it may well be worth bearing in mind that this context change can substantially affect the user’s epistemic situation, particularly if it is perceived as creating a new context and thus severs the link with warranted background

¹² I wish to thank an anonymous reviewer for stressing the importance of this point.

beliefs about reliability. Contextual features must therefore still be weighed along with other features to determine the total evidence available to the user.

I conclude by pointing out some important practical (non-truth-related) features of sources of evidence. In addition to providing a valid link with the truth, evidence of trustworthiness should also be *cognitively simple* and *easy to communicate*. A noteworthy fact about the types of evidence explained above is that they do not rely on difficult-to-process information about the track-record or risks associated with the product. Thus they shift the focus away from traditional conceptions of what counts as relevant evidence for people considering medical treatments. The model of informed consent for surgery and other medical therapies emphasizes information about risk. However, it is very difficult for people to process information about risk. Risk is not cognitively simple or easy to communicate (Fuller et al. 2001; Moore 2008). This has led to the problem that the process of informed consent to medical therapies has mainly a legal or institutional value (protecting the hospital from liability) rather than helping to satisfy the ethical principle of informed consent (Faden and Beauchamp 1986). By broadening the conception of what counts as evidence, this problem can be avoided for DCPIs. There are often evidentially valid and cognitively simple ways to communicate e-trustworthiness and establish sound e-trust in such systems. In this paper I hope to have convinced readers of the importance and feasibility of providing such evidence.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Adler, J. E. (1990). Conservatism and tacit confirmation. *Mind*, 99, 559–570.
- Adler, J. E. (2002). *Belief's own ethics*. Cambridge, MA: MIT Press.
- Anonymous, (2009). The avatar will see you now. *Economist*, 27, 92.
- Anonymous. (2011). *Evaluating health information*. MedlinePlus, US National Library of Medicine. <http://www.nlm.nih.gov/medlineplus/evaluatinghealthinformation.html>. Accessed 26 January 2011.
- Beauchamp, T. L., & Childress, J. F. (2009). *Principles of biomedical ethics* (6th ed.). New York: Oxford University Press.
- Bobylev, V. M., Gromova, E. G., & Toloknov, V. I. (1997). Medical engineering systems for automatic insulin therapy of diabetes. *Biomedical Engineering*, 31(2), 114–116.
- Briggs, P., Burford, B., de Angeli, A., & Lynch, P. (2002). Trust in online advice. *Social Science Computer Review*, 20, 321–332.
- Carusi, A. (2009). Implicit trust in the space of reasons. *Journal of Social Epistemology*, 23, 25–43.
- Case, D. O., Andrews, J. E., Johnson, J. D., & Allard, S. L. (2005). Avoiding versus seeking: The relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. *Journal of the Medical Library Association*, 93, 353–362.
- Coeckelbergh, M. (2010). Health care, capabilities, and AI assistive technology. *Ethical Theory and Moral Practice*, 13, 181–190.
- Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.
- Conee, E., & Feldman, R. (2004). *Evidentialism: Essays in epistemology*. New York: Oxford University Press.
- Ditto, P. H., Hawkins, N. A., & Pizarro, D. A. (2005). Imagining the end of life: On the psychology of advance medical decision making. *Motivation and Emotion*, 29, 481–502.
- Dunn, L. B., Lindamer, L. A., Palmer, B. W., Schneiderman, L. J., & Jeste, D. V. (2001). Enhancing comprehension of consent for research in older patients with psychosis: A randomized study of novel consent procedure. *American Journal of Psychiatry*, 158, 1911–1913.
- Duplaga, M., & Zielinski, K. (2006). Evolution of IT-enhanced healthcare: From telemedicine to e-health. In *Information technology solutions for health care* (pp. 1–21). Dordrecht: Springer.
- Eysenbach, G. (2000). Towards ethical guide guidelines for e-health: JMIR theme issue on eHealth ethics. *Journal of Medical Internet Research*, 2 (1). <http://www.jmir.org/2000/1/e7/>. Accessed 26 January 2011.
- Faden, R., & Beauchamp, T. (1986). *A history and theory of informed consent*. New York: Oxford University Press.
- Fielder, J. (1992). The bioengineer's obligations to patients. *Journal of Investigative Surgery*, 5(3), 201–208.
- Fuller, R., Dudley, N. J., & Blacktop, J. (2001). Risk communication and older people—understanding of probability and risk information by medical inpatients aged 75 years and older. *Age and Ageing*, 30, 473–476.
- Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.
- Hardin, R. (2006). *Trust*. Cambridge, UK: Polity.
- Hieronymi, P. (2005). The wrong kind of reason. *The Journal of Philosophy*, 102, 437–457.
- Hieronymi, P. (2008). The reasons of trust. *Australasian Journal of Philosophy*, 86(2), 213–236.
- Illingworth, P. M. L. (2005). *Trusting medicine*. New York: Routledge.
- Leydon, G. M., Boulton, M., Moynihan, C., Jones, A., Mossman, J., Boudioni, M., et al. (2000). Cancer patients' information needs and information seeking behaviour: In depth interview study. *British Medical Journal*, 320, 909–913.
- Mangunkusumo, R. T., Moorman, P. W., van den Berg-de Ruyter, A. E., van der Lei, Johan., de Koning, H. J., & Raat, H. (2005). Internet-administered adolescent health questionnaires compared with a paper version in a randomized study. *Journal of Adolescent Health*, 36 (1), 70.e1–70.e6.
- Möllering, G. (2006). *Trust: Reason, routine, reflexivity*. Amsterdam: Elsevier.
- Moore, R. A. (2008). What do we know about communicating risk? A brief review and suggestion for contextualising serious, but rare, risk, and the example of cox-2 selective and non-selective NSAIDs. *Arthritis Research & Therapy*, 10, R20.
- Myerburg, R. J., Feigal, D. W., Jr., & Lindsay, B. D. (2006). Life-threatening malfunction of implantable cardiac devices. *New England Journal of Medicine*, 354, 2309–2311.
- National commission for the protection of human subjects of biomedical and behavioral research. (1979). *The Belmont report*. US Department of Health and Human Services. <http://www.hhs.gov/ohrp/policy/belmont.html>. Accessed 27 January 2011.
- Nickel, P. J. (2001). Moral testimony and its authority. *Ethical Theory and Moral Practice*, 4, 253–266.
- Nickel, P. J. (2012). Trust in technological systems. In M. J. de Vries, S. O. Hansson, & A. W. M. Meijers. *Norms and the artificial: Moral and non-moral norms in technology*. Dordrecht: Springer.
- O'Neill, O. (2002). *Autonomy and trust in bioethics*. New York: Cambridge University Press.

- Østbye, T., Yarnall, K. S. H., Krause, K. M., Pollak, K. I., Gradison, M., & Michener, J. L. (2005). Is there time for management of patients with chronic diseases in primary care? *Annals of Family Medicine*, 3, 209–214.
- Pettit, P. (1995). The cunning of trust. *Philosophy & Public Affairs*, 24, 202–225.
- Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Selmi, P. M., Klein, M. H., Greist, J. H., Sorrell, S. P., & Erdman, H. P. (1991). Computer-administered therapy for depression. *MD Computing*, 8(2), 98–102.
- Stanberry, B. (2001). Legal ethical and risk issues in telemedicine. *Computer Methods and Programs in Biomedicine*, 64(3), 225–233.
- Taddeo, M. (2009). Defining trust and e-trust: Old theories and new problems. *International Journal of Technology and Human Interaction*, 5, 23–35.
- Thomson, J. J. (1971). A defense of abortion. *Philosophy & Public Affairs*, 1(1), 47–66.
- Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, 102, 269–283.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.
- Vallero, D. (2007). *Biomedical ethics for engineers: Ethics and decision making in biomedical and biosystem engineering*. London: Academic Press.
- van Uden-Kraan, C. F., Drossaert, H. C., Taal, E., Smit, W. M., Bernelet Moens, H. J., Siesling, S., et al. (2009). Health-related internet use by patients with somatic diseases: Frequency of use and characteristics of users. *Informatics for Health and Social Care*, 34(1), 18–29.
- Walker, M. U. (2006). *Moral repair: Reconstructing moral relations after wrongdoing*. New York: Cambridge University Press.
- Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 345–411). New York: Elsevier.