



Niṣkāmakarma and the Prisoner's Dilemma

Tommi Lehtonen¹ 

Published online: 14 July 2020
© The Author(s) 2020

Abstract

The *Bhagavadgītā*, part of the sixth book of the Hindu epic The *Mahābhārata*, offers a practical approach to *mokṣa*, or liberation, and freedom from *saṃsāra*, or the cycle of death and rebirth. According to the approach, known as *karmayoga* ('the yoga of action'), salvation results from attention to duty and the recognition of past acts that inform the present and will direct the future. In the *Bhagavadgītā*, Kṛṣṇa advocates selfless action as the ideal path to realizing the truth about oneself as well as the ultimate reality. Kṛṣṇa proclaims that humans have rights only to actions and not to their results, whether good or bad (2.47). Therefore, humans should not desire any results whatsoever. The prisoner's dilemma is a fictional story that shows why individuals who seek only their personal benefit meet worse outcomes than those possible by cooperating with others. The dilemma provides an effective, albeit often overlooked, method for studying the Hindu principle of *niṣkāmakarma* ('desireless action') that is arguably the central teaching of the *Bhagavadgītā*. In the context of the prisoner's dilemma, a prisoner who wants to uphold *niṣkāmakarma* may choose one of two decision-making strategies: to be indifferent and leave the decision to chance or to either pursue the common good or the other person's benefit instead of his or her own. Assuming that followers of *niṣkāmakarma* can be goal-oriented, the second strategy is more appropriate than the first, as long as one pursues unselfish goals and remains both indifferent and uncommitted to personal benefit.

Keywords *Niṣkāmakarma* · The prisoner's dilemma · Desireless action · Benefit · Selflessness · The *Bhagavad Gītā*

Introduction

This paper aims to shed new light on the topic of *niṣkāmakarma* ('desireless action') in the *Bhagavadgītā*. Specifically, it does so through examination of the thought

✉ Tommi Lehtonen
tommi.lehtonen@univaasa.fi

¹ School of Management, University of Vaasa, PO Box 700, FI-65101 Vaasa, Finland

experiment known as the ‘prisoner’s dilemma’. The dilemma shows why two completely rational individuals might not cooperate even when doing so appears to be in their best interests. The dilemma provides a well-formulated, theoretical game context for developing an understanding of *niṣkāmakarma*. In the context of the prisoner’s dilemma, a prisoner who wants to uphold the principle should choose one of two decision-making strategies: to pursue the common good or to pursue the other person’s benefit instead of his or her own. Assuming that followers of *niṣkāmakarma* can be goal-oriented, the first strategy is no less appropriate than the second, as long as one remains both indifferent and uncommitted to personal benefit.

However, things become more complicated when the notion of liberation in the *Bhagavadgītā* (hereafter The *Gītā*) is considered. The *Gītā* advises human beings to seek liberation (*mokṣa*) as the final goal of life (2.64–66, 2.70–72, 4.19–23, 6.15, 6.36, 17.25). One may wonder whether the concern for liberation contradicts selflessness and desirelessness: Why would one be concerned about liberation unless one thinks of oneself as something whose liberation is desirable? The *Gītā* does not provide a straightforward answer to this question. Instead, it (3.20, 3.25, 18.5–7) argues that contributing to the common good dutifully and selflessly is a proper way to attain liberation (Sreekumar 2012, pp. 277–278, pp. 308–310). From this the following dilemma arises, one should strive for liberation while becoming liberated from self-interested strivings. Here, the notion of *niṣkāmakarma* comes in, simultaneously mitigating and exacerbating the dilemma: The way to liberation is through desireless action. How should such an ambiguous notion as desireless action be understood?

As I elaborate in this paper, the new viewpoint provided by the ‘prisoner’s dilemma’ can enhance current understandings of selflessness and desirelessness, as well as fill a gap in research on the principle of *niṣkāmakarma*. First, the prisoner’s dilemma presents and illuminates the two basic strategies of decision-making: focusing on self-interest versus pursuing the greater good. Moreover, the prisoner’s dilemma suggests that those strategies do not necessarily go hand in hand, which is exactly what Kṛṣṇa (indirectly) points out in the *Bhagavadgītā* when he warns against acting for personal gain (2.47–48, 6.1). Most importantly, the prisoner’s dilemma reveals the rationality of striving to advance the common good, which is a prominent teaching in the *Bhagavadgītā* (3.19–20, 3.25) and a move away from self-interest. Before elaborating upon the merits of the viewpoint, however, I make some general remarks on the concept of *niṣkāmakarma* and discuss a philosophically relevant aspect of translation related to the outcomes of desireless action.

The Concept of *Niṣkāmakarma*

The *Bhagavadgītā*, part of the sixth book of the Hindu epic The *Mahābhārata* (common practice), offers a practical approach to *mokṣa*, or liberation, and freedom from *saṃsāra*, or the cycle of death and rebirth. According to the approach, known as *karmayoga* (‘the yoga of action’), salvation results from attention to duty and the recognition of past acts that inform the present and will direct the future (Singh 1991, p. 107). In the *Gītā*, Kṛṣṇa advocates selfless action as the ideal path to realizing the truth about oneself as well as the ultimate reality. According to Kṛṣṇa, action taken without self-centered expectations or consideration of outcomes tends to purify one’s mind.

Moreover, action without desire gradually makes an individual fit to see the value of mind control and the benefits of renouncing the action itself (2.41, 2.48–49, 6.1–4, 6.7, 6.24–27). The benefits of such renunciation essentially include liberation from an attachment to worldly bonds and suffering.

The central tenet of the *karmayoga* path to liberation is the principle of *niṣkāmakarma*. *Niṣkāmakarma* refers to an action performed without any expectation of reward or result whatsoever. The *Bhagavadgītā* calls that idea 'inaction in action and action in inaction' (4.18). Even if the term does not appear in the *Bhagavadgītā* itself (Fowler 2012, p. xliii–iv), *niṣkāmakarma* is arguably the central teaching of the text (Chakraborty 1996, 1998). At the text level, this argument is based on verse 18.6:

But having relinquished [all] attachment and [actions'] fruits, even these actions should be performed—this is My decided ultimate (*uttama*) conviction (*mata*), O son-of-Prithā (Feuerstein 2014, 299).

However, one must be on guard against uncritically adopting the term *niṣkāmakarma* as shorthand for the central teaching of the *Bhagavadgītā*. The danger lies, first, in the fact that *kāma* is not as broad in its semantic signification as the English desire. Although there can be pure, Sattvic 'desires' such as the desire for liberation (18.26, 18.30), it would not usually be termed *kāma* by pre-modern Sanskrit authors (i.e., *mumuṣā* is not *kāma*). Moreover, the *Bhagavadgītā* is in many ways a difficult document to interpret, not least because it deals with a wide variety of topics that relate not only to ethics but also to metaphysics, epistemology, eschatology, soteriology, yogic technique, and so on (Sreekumar 2012, pp. 279–280). Therefore, it remains open to question—and depends on the point of view—what the central teaching of the *Bhagavadgītā* is.

It is well known that a basic belief in Hinduism is the law of karma or action—briefly, that every good thought, word, or deed begets a similarly good reaction in this or in a subsequent life and, by the same token, that every unkind thought or evil deed ultimately comes back to harm the actor in this life or the next. In traditional thought, karma consists of three stages. First, *prārabdha karma* refers to the accumulated effects of past deeds that one experiences in his or her present life. Second, *sañcitakarma* refers to acts performed either in this life or in a previous one but that have not yet begun to bear fruit. Third and last, *āgāmikarma* refers to acts performed in this life, the results of which are yet to come. *Niṣkāmakarma* is considered to be the ideal kind of karma, for it is obtainable only by not seeking any reward (Singh 1991, p. 120).

Different Interpretations of *Niṣkāmakarma*

In the second chapter of the *Bhagavadgītā*, Kṛṣṇa proclaims that humans have rights only to actions and not to their results, whether good or bad (2.47–48). Therefore, they should not desire any results. Of utmost importance in such thinking is that humans are not doomed to idleness or inactivity but instead should act according to the requirements of morality and decency. At the same time, they are not entitled to the results of their actions and, as such, should not be selfishly concerned about them. Although admirable to perform good deeds and participate in well-intentioned activities, the

outcomes of those actions should not be associated with their actors. Instead, according to the *Bhagavadgītā*, the positive outcomes of actions should be understood to represent the common good (3.19–20, 3.25), which should be accessible to every member of society. Thus, the outcomes of actions are not ultimately under the control of humans, who are, after all, instruments of the becoming of the ultimate reality, often understood as the fulfilment of God's eternal designs (*Bhagavadgītā* 11.15–34; Singh 1991, p. 107). Thus, humans have the right to use the outcomes of their actions for good, but they do not own those outcomes.

When reading Kṛṣṇa's teaching—in short, that people should not desire results—with sensitivity to philosophy, the advice seems problematic and strange, for an action and its results are internally related. For example, if I manage to open a window, which can be opened only by my action, then the window's being open is directly due to my action (Kim 1993, p. 26).

Of course, interpretations of Kṛṣṇa's meaning are inevitably subject to the loose and ambiguous use of language in translation. In the *Bhagavadgītā* (2.47), the Sanskrit term for result is *phala*, or 'fruit', which many translators replace with either 'result' or 'reward'. In that context, the term reward relates to the idea of the universal moral bookkeeping of karmic merits as the basis for rewards. Clearly, the *Bhagavadgītā* thus distinguishes an action from its outcomes, whether those outcomes are called 'results', 'fruit', or 'rewards'. According to the *Bhagavadgītā*, performing a deed is therefore separate from its result. Nevertheless, a possible interpretation is that the *Bhagavadgītā*, in using the term *phala* vaguely, conflates the results of an action with its consequences. Indeed, along with its results, an action can have intended or unintended consequences. Whereas the relationship between an action and its result is intrinsic, the relationship between an action and its consequences is extrinsic or causal (von Wright 1963; Raz 1975).

In an elaboration upon the difference between the results and consequences of an action, Georg Henrik von Wright (1963, p. 39) has explained:

By the *result* of an act we can understand either the change corresponding to this act or alternatively the end-state [. . .] of this change. Thus by the result of the act of opening a certain window we can understand either the fact that the window is opening (changes from closed to open) or the fact that it is open. On either way of understanding the notion of a result of action the tie between the act and its result is intrinsic. [. . .]. Unlike the relation between an act and its result the relation between an act and its consequences is extrinsic (causal).

To reveal how complex the issue can be, von Wright has added that

One and the same change or state of affairs can be both the result and a consequence of an action. What makes it the one or the other depends upon the agent's intention in acting and upon other circumstances. (p. 40)

According to von Wright's analysis, the agent's intention determines whether a state of affairs is the result or the consequence of an action. The consequences of opening a window can include fresh air, and a fly coming into the room. Fresh air is often an intended consequence, whereas a fly is more likely an unintended one. Moreover, one

can reinterpret opening a window as the act of letting in fresh air, in which case fresh air is the result of the act.

Based on those initial considerations, Kṛṣṇa's advice that people should not desire any results from their actions can be interpreted in different ways (Sreekumar 2012, p. 302). One might interpret it as tongue in cheek, as Kṛṣṇa's advising people to take action but to avoid exerting much effort in the process. That interpretation clearly differs from one renouncing action or, somewhat differently, endorsing inaction (Chakrabarti 1983). Could the interpretation—that people should perform deeds without striving or taking the trouble to make those deeds matter—be correct? There is no reason to think so, for Kṛṣṇa does not also advise carelessness or sluggishness or leaving deeds unfinished. Instead, he suggests focusing not on personal benefits as outcomes of actions but on duties and the common good, as I soon demonstrate. Textual evidence for the latter interpretation abounds, as the following examples show:

In action alone is your rightful-interest (*adhikāra*), never in [its] fruit. Let not your motive be the fruit of action; nor let your attachment be to inaction (*akarman*). Steadfast in Yoga, perform actions abandoning attachment, O Dhanamjaya, [always] remaining the same in success and failure. Yoga is called equanimity. (2.47–48; Feuerstein 2014, 107, 109)

Renouncing all actions in Me, with the mind [turned toward] the basis-of-self, [and] having become without hope, without [the sense of] “mine,” [with your] fever-of-anxiety departed—fight! (3.30; Feuerstein 2014, 129)

[He who is] yoked, having relinquished the fruit of action, attains ultimate peace. The unyoked [individual], acting from desire and attached to the fruit [of action], is bound [by karma] (5.12; Feuerstein 2014, 151)

He who performs the action to be done, regardless of action's fruit, is a renouncer and a *yogin*; not [so is he who is] without the [sacrificial] fire and is inactive. (6.1; Feuerstein 2014, 157)

[When] he performs necessary action that is indeed to be done, O Arjuna, and by relinquishing attachment and the fruit [of one's action]—the relinquishment is deemed [to be] *sattva*-natured. (18.9; Feuerstein 2014, 301)

Moreover, Kṛṣṇa proclaims that performing actions without entangling them with desires will afford release from the chains of rebirth (*Bhagavadgītā* 2.51; 4.18–20).

From an alternative perspective, Brodbeck (2004, pp. 84, 89, 92–93, 95, 98, 100) has sharply criticized any attempts to interpret Kṛṣṇa's advice positively and claimed instead that Kṛṣṇa is bluffing. According to Brodbeck, Kṛṣṇa takes a stance advocating determinism: that is, the view that all which happens occurs necessarily and that nothing may happen in any other way. With that stance, he fools Arjuna, the commander of the Pandava army, into believing that he can and should decide to fight, even if such a decision is impossible; after all, Arjuna is not free to choose but acts according to the force of destiny. In Brodbeck's view, Kṛṣṇa cleverly does not try to convince

Arjuna of the reality of determinism but instead appeals to Arjuna's duties as a soldier. However, that interpretation, though intriguing, is unconvincing, as well as controversial, given the *Bhagavadgītā*'s general position on desire and liberation. If everything, including opinions and intentions, is determined, then being concerned about changing oneself would be as pointless as being concerned about the impossibility of such change. Whereas such self-change, if possible, would be highly advisable, if it is impossible, then entertaining it is pointless. At the same time, if you are determined either to entertain the thought of such change or to fear the consequences of not changing, then you cannot but consider changing. However, textual evidence justifying the assumption that the *Bhagavadgītā* advocates such a complex perspective on desires and liberation is thin and ambiguous (see 3.5, 18.58–60).

It seems more plausible, as several other authors have pointed out, to interpret Kṛṣṇa's advice to mean that people should perform deeds out of a sense of duty, not for self-centered gain (*Bhagavadgītā* 3.19; 18.9). This interpretation conveys a certain idea of humbleness and serenity: humans should not take the outcomes of their actions as rights or something that they are entitled to but should accept positive outcomes as (undeserved) gifts and should not complain about negative outcomes. Even if the *Gītā* thus advises not focusing on personal benefit, some deeds—for example, ones related to fulfilling basic physical needs such as eating and sleeping—are necessarily self-centered, such that performing them inevitably responds to the self. Arguably, in fulfilling those basic needs, a healthy self-interest is necessary and thus does not contradict duty.

Regarding the *Bhagavadgītā*'s concept of the person, *ātman* ('the true, inmost self')—characterized in *mahāvākyas*, or the great sayings of the *Upaniṣads*, to be identical with *Brahman* or the ultimate reality—is not the doer of deeds but an observer and adviser similar to Kṛṣṇa, the charioteer to Prince Arjuna. In the *Bhagavadgītā*, Arjuna, shocked at the thought of fighting his Kaurava relatives, hesitates before the battle at Kurukshetra (1.26–2.9). Although Arjuna orders his charioteer to withdraw, Kṛṣṇa urges Arjuna to ready himself for battle and to regard pleasure and pain, gain and loss, and victory and failure all as one and the same. Only when Arjuna has renounced interest in the fruits of his actions does he find true peace (2.14–38). Traditionally, readers have understood Kṛṣṇa and Arjuna's discussion to imply that the doer of deeds is *jīva*, or the empirical or seeming self, equipped with a mind and body. Moreover, according to the *Bhagavadgītā*, salvation ultimately derives from the recognition that the true self is not *jīva*, the doer, and thus that the true self does not reap the fruits of action, either (2.47). Whereas the true or transcendental (or second order) self, *ātman*, is concerned with fulfilling moral duties (2.31–39) and with observing and evaluating the actions of the empirical self, the empirical self is concerned with obtaining benefits and the results of actions (2.2–8). Consequently, salvation is not only an ethical topic in the *Bhagavadgītā* but also a serious onto-epistemological one, for a true understanding of the self and of ultimate reality is intrinsic to salvation.

In fewer words, the *Bhagavadgītā* teaches that people should fulfil their duties and advance the common good, or *lokasaṅgraha* ('the holding-together of the world', 'the welfare of the world', 3.19–20, 3.25). At the same time, they should understand that *ātman*, or the inmost self, is not the doer of deeds but an observer and adviser. As Indian social reformer and independence activist Bal Gangadhar Tilak (1936, 466) has worded the matter:

A man should not entertain the proud or desirous thought that “I shall bring about *lokasaṅgraha*” [. . .]. A man has to bring about *lokasaṅgraha* merely as a duty.

To return to Kṛṣṇa's advice that people should not desire results, von Wright has provided the means for another interpretation, according to which individuals, when acting, should focus on changes corresponding to the actions, not their ultimate effect. A suitable metaphor for that interpretation is the Zen archery competition at which a master advises his disciples to focus their attention on aiming, not on scoring (Herrigel 1999). Similarly, the controlled performance of a gymnastics routine is arguably more important than the points given by a jury. In that context, the Western Aristotelian distinction between doing and making (or between action and production) can be illuminating. Gymnastics, as well as dance and music, are examples of Aristotelian *praxis* ('action'), the result of which is not separate from but included in the activity. Making and producing, or *poiesis*, in turn, is an activity in which the result is separate from the activity, as in house building and thesis writing (*Nichomachean Ethics* I.1, 1094a1–5, VI.4, 1140a1–24, VI.5, 1140b6–7, 1140b7; Aristotle 1999, 1, 88–89). By extension, Kṛṣṇa's advice can be interpreted to encourage concentration on doing or performing and not the outcome or end-state of those activities. That solution, however, is problematic, especially in the context of production, because it diminishes the value of the result (e.g., profit) that is fundamental in, for example, financial performance.

Situations in which one takes action without considering its outcomes include:

1. When one does not wait for anything related to acting but simply acts (e.g., out of duty or gratitude);
2. When one is in a state of flow or so deeply immersed in an activity that he or she performs an action without thinking about its results or consequences; and
3. When one acts like an automaton and does not know or understand what he or she is doing.

The first two—not waiting for anything before acting and acting in a state of flow—are more plausible interpretations of Kṛṣṇa's advice, for he does not make any reference to acting unconsciously.

A relevant but seemingly contradictory question is: In what way, and to what extent, can *niṣkāmakarma* benefit agents? In answering that question, I should clarify that, based on the foregoing analysis, desireless does not mean purposeless, indifferent, or unintentional, but dutiful, conscientiously or obediently fulfilling one's duty or being motivated by duty instead of by the potential of reward. Moreover, because duty implies obligation, fulfilling one's duty is, by definition, obligatory and necessary. In terms of psychological benefits, acting without desire can produce byproducts such as freedom from stress, a relaxed performance and life with fewer disappointments, all because expectations are not directed towards success or failure but towards the fulfilment of duty. The potential risks of acting without desire, in turn, include negligence, passivity, and too much relaxation.

Based on the aforesaid, Kṛṣṇa's advice that people should not desire any results from their actions should be interpreted either as an obligation to act out of duty and unselfishly, for which a suitable motto might be ‘True love is giving, not receiving’, or as the wisdom to concentrate on doing an activity, not on its outcome. At the same

time, both interpretations are compatible and not exclusive; people can both act altruistically and concentrate on performing activities.

In reference to the notion that *niṣkāmakarma* can be acquired only by renouncing it, a relevant question is whether *niṣkāmakarma* is an ability or a disposition that can be acquired. If the latter is true, then another question is whether *niṣkāmakarma* requires or presupposes free will. One can have free will in at least three senses: as the necessary condition for moral responsibility (McKenna and Pereboom 2016, p. 8), as the ability to do other than what one has done (van Inwagen 2017) and as the ability to bring about one's characteristics, virtues and strengths by action (Kane 1996, p. 4). By extension, three interpretations of *niṣkāmakarma* can be distinguished: the disposition of dutifulness, the disposition to advance the common good and the disposition to follow the second-order will, also known as the observer–adviser concept of the self. Based on what I have shown, *niṣkāmakarma* requires acting either like a robot without considering the results of one's actions or like a virtuous person who has developed a character not oriented towards achieving the results of his or her actions. When so understood, *niṣkāmakarma* requires free will, at least in the context of the third sense (i.e., the ability to bring about one's virtues and strengths). At the same time, by emphasizing the importance of fulfilling one's duties, *niṣkāmakarma* seems to uphold the first sense of free will, for it would be pointless to exhort an automaton to fulfil its duties and responsibilities. The principle of *niṣkāmakarma* also presupposes that, in deeds and actions, one can focus on the results or on the performance and thereby do otherwise than what one has done. Thus, all three meanings of free will are relevant to and accessible in the concept of *niṣkāmakarma*.

Despite a plausible understanding of how to interpret Kṛṣṇa's advice—that people should not focus on results when engaging into action, which however does not rule out the possibility of taking into consideration consequences—the prisoner's dilemma provides a means to introduce a fourth interpretation: to be indifferent when choosing among a range of actions and to leave the decision to chance, fate, or providence (Brodbeck 2004). Although that interpretation is doubtful, because it has a very thin and ambiguous evidence base in the *Bhagavadgītā* (2.32, 4.22) and ignores the focus on duties central to Kṛṣṇa's advice to Arjuna (3.19, 18.9), to better understand the role of chance in decision-making and desireless action, the prisoner's dilemma warrants consideration. However, this is not the only, or even the main, reason why the dilemma is worth considering. Most importantly, the dilemma reveals the rationality of striving to advance the common good, i.e., seeing one's own well-being as connected to the larger well-being of society and of fellow human beings. As was seen, this is a prominent teaching in the *Bhagavadgītā*.

The Prisoner's Dilemma

Formulated in 1950 by Merrill Flood and Melvin Dresher and later formalized by Albert Tucker (Peterson 2009; Poundstone 1992), the prisoner's dilemma is a fictional story of a decision-making situation in which individuals seeking their own benefit each end up with a worse outcome than what they could have achieved by cooperating. Used to demonstrating decision-making in the face of risk and uncertainty, the dilemma

has inspired numerous discussions and various analyses among game and decision theorists.

In the story, John and Mary, the accomplices to a crime, have been apprehended and threatened with imprisonment for their wrongdoing. Because neither is able to communicate with the other, they can remain silent or they can expose each other. If both remain silent, then each will receive a 1-year prison sentence. However, if one of them confesses, then he or she will go free while the accomplice is jailed for three years. If both confess, then each will receive a 2-year sentence. Furthermore, both of them are aware of the consequences of all three options (cf. Peterson 2009).

The true dilemma of the situation is that, whatever one chooses, the other had better confess, even though the common good would be for both of them to remain silent. Written as two Nash equilibria, the dilemma implies that neither prisoner has anything to gain by changing his or her strategy alone. The situation can be conceived as shown in Table 1, in which the left-hand number indicates the prison sentence, in years that John will receive, whereas the right-hand number indicates the sentence, also in years that Mary will receive. If both accomplices confess, then the sentence for each will be 2 years; however, if both remain silent, then each will receive a sentence of only a year.

If John remains silent, then Mary will be sentenced to a year in prison if she also remains silent. However, if Mary exposes John, then she will be set free. In another scenario, if John snitches on Mary, then Mary, if she remains silent, will be sentenced to three years but, if she confesses, only to two. Thus, the prisoner who acts in the interest of his or her best interest would snitch on the other and thereby ensure a shorter sentence for him- or herself, instead of remaining silent and inevitably receiving a sentence. Although John and Mary would act rationally by seeking his or her own best benefit, as a twosome, they face a worse outcome by acting selfishly than they would by cooperating (Myerson 1991; Peterson 2009).

The collectively suboptimal outcome depends on either the absence of an enforceable agreement or intrinsic trust between the prisoners and a lack of information about each other's intentions. Rationality and self-interest would force each prisoner to betray the other and thus choose an outcome worse for both of them than the outcome afforded were they to cooperate and thereby minimize their total number of years in prison. The dilemma showcases that what is optimal for each risk-averse individual needs not coincide with what is collectively optimal (Peterson 2009). As mentioned, the prisoner's dilemma also presents and illuminates the two basic strategies of decision-making: focusing on self-interest versus pursuing the greater good. Moreover, it suggests that those strategies are not necessarily associated, which is exactly what Kṛṣṇa indicates in the *Bhagavadgītā* when he warns against acting for personal gain (2.47–48, 6.1).

Table 1 Strategies and outcomes available to John and Mary in the prisoner's dilemma

	Mary confesses	Mary remains silent
John confesses	2, 2	0, 3
John remains silent	3, 0	1, 1

Of course, it is possible to defend an even more altruistic perspective on the core of the prisoner's dilemma. According to that perspective, the only genuinely altruistic strategy to achieve the combination of a short sentence and help one's accomplice is to remain silent, even when faced with the risk that the accomplice may confess. After all, altruism requires accepting the risk that one's self-interest might remain unfulfilled. From the alternative perspective of moral theory, both act utilitarians as well as rule utilitarians would theoretically support the view that both prisoners should remain silent because, by doing so, the greatest good for the greatest number would be achieved in the particular case (i.e., act utilitarianism) and in general (i.e., rule utilitarianism), provided that no one other than John and Mary are involved. Kantian deontologists, by extension, would likely argue that both prisoners ought to tell the truth and confess, because telling the truth is a duty, Kant says, and not an action to be performed on a case-by-case basis (*Groundwork of the Metaphysics of Morals*, 4:389, 4:403, 4:441; Kant 2002, 5, 19, 58–59). That said, an arbitrary but nevertheless notable feature of the dilemma concerns prisoners' liability to confess and remain silent. Even without a pro tanto or prima facie duty to tell the truth and despite prison being the story's context, the prisoner's dilemma is more deeply about the conflict between individual benefit and mutual advantage (Weirich 2015).

In the prisoner's dilemma, a prisoner who wants to follow the principle of desireless action would theoretically follow one of two decision-making strategies: be indifferent and leave the decision to chance (e.g., by arbitrarily drawing lots) or either pursue the common good or the other prisoner's benefit (e.g., by remaining silent) instead of seeking to benefit him- or herself (e.g., by snitching). Not only is the second strategy ethically more appropriate, but it also follows Kṛṣṇa's advice in the *Gītā*, at least assuming that individuals who want to act without desire (i.e., followers of *niṣkāmakarma*) can be goal-oriented, pursue unselfish goals, and remain both indifferent and uncommitted to personal benefit. Readers such as Chuang (2015) have espoused that interpretation of Kṛṣṇa's advice and dubbed it 'benevolent action'. I agree with this interpretation and aim to expand upon it by suggesting, below, a particular understanding of Sreekumar's (2012) argument that the *Bhagavadgītā* advances consequentialist ethics. In contrast, Framarin (2009) has effectively problematized prioritizing unselfish desires. He has observed that contemporary readers almost unanimously contend that, since all action is motivated by desire, desireless action is an oxymoron; consequently such action is performed without selfish desire, meaning that unselfish desire is permissible. Framarin has also pointed out, however, that arguments for that view are unconvincing, since the doctrine of desireless action should be taken literally: as advice not to act without any desire at all (Chakrabarti 1983). Following such a doctrine would require people to act as automatons without wanting or waiting for anything before acting, which, from a psychological standpoint and based on the text evidence in the *Bhagavadgītā*, is a highly unintuitive and implausible perspective. Therefore, I think that Framarin's argument is not convincing.

According to a common view, the *Bhagavadgītā* advances duty-based ethics through the mouth of Kṛṣṇa, whereas Arjuna appeals to consequentialist considerations. This view is represented, for example, by Sen (2009, pp. 23–24, 208–217), who emphasizes that the *Bhagavadgītā* should be seen as a classic debate between deontological and consequential ethics. Contrary to this view, Sreekumar (2012, pp. 278, 299–300, 303–310) argues that the *Bhagavadgītā* advances consequentialism

rather than duty-based ethics. Sreekumar forcefully defends the view that Krishna's ethical thinking is a distinctive kind of rule-consequentialism that takes as intrinsically valuable the twin consequences of liberation (*mokṣa*) and the common good (*lokasaṅgraha*). According to Sreekumar (2012, p. 301), the core of the doctrine of *niṣkāmakarma* is that the agent must detach himself or herself from all those consequences of action that redound directly or indirectly to his or her personal advantage, egoistically considered. I agree with Sreekumar.

The discussed interpretations of Kṛṣṇa's advice are listed in Table 2.

Four of these interpretations are relevant based on the text evidence in the *Bhagavadgītā*. The relevant interpretations are 1, 2, 3, and 5. Interpretation 1 is the most traditional, while interpretation 2 is supported both by the text evidence in the *Gītā* and by the prisoner's dilemma. Interpretation 3 is a kind of onto-epistemological extension of both interpretations 1 and 2: a right understanding of the true self involves selfless dutifulness and serving the common good. Interpretation 5 directs the attention to the actual performance of the deed rather than to the result of the deed, which can be matched both with interpretations 1 and 2. The compatibility of interpretations 1 and 2 requires that contributing to the common good is understood as an absolute duty. Moreover, even though interpretation 5 rejects the result of action as the focus of the agent, it does not prevent the agent from focusing on the consequences of his or her action. This represents a consequentialist argument compatible with interpretation 2. Interpretations 4 and 6 are more or less irrelevant for the *Bhagavadgītā* and therefore are not discussed further.

Sreekumar (2012, p. 307, n. 75) argues that no ethical theory can be both deontological and consequentialist. However, he adds that an ethical theory can be rule-consequentialist. What often remains unobserved—and this also pertains to Sreekumar's view—is that different rule-based accounts of ethics, like consequentialism and deontological ethics, can be, and in everyday life often are, integrated in a hierarchical order and are used simultaneously for different purposes, such as moral evaluation and education, encouragement, and disapproval. The levels of a norm-hierarchy can be expressed as definitory rules and strategic rules. Definitory rules define the basic moves, i.e., what is and what is not admissible in the game, and strategic rules explain how to play the game (Hintikka 1999, p. 98). Consequentialism and deontological ethics need not contradict each other in so far as they operate at different levels of a rule hierarchy, i.e., for different purposes, and approach human action from different angles. Specifically, these different purposes are to fulfil duties and responsibilities, on the one hand, and to avoid harms and to pursue benefits, on the

Table 2 Different interpretations of Kṛṣṇa's advice

-
1. Perform deeds out of duty.
 2. Pursue the common good or the benefit of the other instead of personal benefit.
 3. Know and follow the true self, *ātman*.
 4. Be indifferent when choosing among actions and leave the decision to chance, fate or providence.
 5. Focus on acting rather than on its outcomes.
 6. Perform deeds without taking the trouble to accomplish them.
-

other. I want to emphasize that this ‘rules of different levels’ interpretation is not a magic trick that would eliminate the major difference between consequentialism and deontological ethics, namely that in deontological ethics the justification of moral duties is independent of the consequences of performing them, while in consequentialism the moral duties are justified in terms of the good consequences that the performance of those duties will have. The ‘rules of different levels’ concept also provides a solution to the dilemma mentioned in the beginning of this essay: The striving for liberation is the definitory goal while renouncing desires is a strategic mean to reach that goal. The fundamental difference between consequentialism and deontological ethics is thus admitted, but it must be added that this difference does not prevent a certain kind of reconciliation: The rules of different levels (i.e., the level of definitory rules and the level of strategic rules) can be coordinated such that deontological duties limit consequentialist duties. For example, the fundamental human rights and related prohibitions of punishing the innocent and mistreating anyone limit the utilitarian use of punishment as a deterrent and regulate the treatment of minorities. Moreover, consequentialism and deontological ethics often require the same action but for different reasons. For example, telling the truth and keeping one’s promises are plausible deontological requirements as well as the necessary conditions for viable societies and economies. Thus, the same rule such as ‘one should tell the truth’ or ‘one should keep one’s promises’ can be arrived at in different ways and from different reasons and has been conceived from different perspectives by deontologists and consequentialists alike.

One may consider it natural that the definitory rules of ethics are deontological while strategic rules can be consequentialist: Deontological ethics defines actions that are permitted or absolutely required or prohibited, while consequentialism defines how to maximize the benefit or to minimize the suffering. One can add that the role of strategic rules is suitable for deontological ethics that does not provide goals for human life but rules for any situation. However, also in regard to the maximizing of benefit and minimizing of harm, actions are allowed only if they are required or permitted by definitory rules. Referring to Sreekumar’s interpretation of the *Bhagavadgītā*, seeking for liberation is a consequentialist goal whose realization requires that the duty to contribute to the common good is met (Sreekumar 2012, 308, 309). Thus, one can conclude that in the *Gītā*, definitory rules of the ‘game of life’ are consequentialist (e.g., the rule that liberation is to be sought), while strategic rules can be both deontological (e.g., the rule that liberation requires fulfilling universal moral obligations) and consequentialist (e.g., the rule that liberation requires doing one’s professional responsibilities, like the duties of a soldier, and contributing to the common good).

To sum the above discussion, the desirelessness of an action, in the context of the prisoner’s dilemma, has to be addressed in relation to the following options:

- A. Acting out of self-interest (i.e., in opposition to the common good) and thereby either maximizing one’s personal benefit or minimizing harm to oneself; or
- B. Acting collaboratively (i.e., generously towards others) or altruistically (i.e., benevolently and with concern for the wellbeing of others) and thereby maximizing either the collective benefit or the benefit of others.

Confessing and remaining silent are the concrete actions by which those options can be realized in the prisoner's dilemma. An indicator of indifference, by contrast, would be that neither of the given options is considered to be eligible and that neither is sought.

According to the first interpretation of *niṣkāmakarma* in light of the prisoner's dilemma, action without desire means that one's personal benefit, the collective benefit, and the benefit of others are not goals. However, if the choice has to be made, then an indifferent means or one free of desire is to either draw lots or leave the decision to chance. Clearly, neither the *Bhagavadgītā* nor the principle of *niṣkāmakarma* recommends making decisions of human action by resorting to chance. Thus, the *Bhagavadgītā* does not recommend the freedom of indifference or arbitrary randomness but exhorts people to act according to duty (3.19, 18.9).

The question remains, however, whether another way for an action to be desireless is possible. In response, with reference to the principle of *niṣkāmakarma*, people should choose the second option and act collaboratively and altruistically (i.e., with concern for the wellbeing of others) instead of acting out of self-interest. Such an answer means that, even if people should not aim to receive the fruits of their actions for themselves, they should at least aim for the common good and the benefit of and positive consequences for others. Advice from chapter 3 of the *Bhagavadgītā* reiterates that answer clearly: 'Even considering only the world's welfare, you ought to act' (3.20; Feuerstein 2014, p. 125) and 'the wise should act thus unattached, desiring to accomplish the world's welfare' (3.25; Feuerstein 2014, p. 127). In terms of options listed in Table 2, the solution approximates what is known as 'benevolent action' (option 2) (Chuang 2015).

Nevertheless, it is still doubtful whether an altruistic or collaborative motive for action truly represents desireless or disinterested action. One answer is that it depends on the point of view. From a moral point of view, an altruistic action is arguably desireless, because, by definition, it is not based on a desire for one's personal benefit. However, all action requires intention, as Davidson (1980) has pointed out, and altruistic or collaborative action is no exception. If one's intention is to maximize the collective benefit or the benefit of another, and if the person knows which results or end-state could make his or her intention true, then he or she intends to perform an action that is not desireless—at least not in any absolute sense—but deliberate and purposeful. This is an example of the danger of adopting the term *niṣkāmakarma* as shorthand for the central teaching of the *Bhagavadgītā*, when in fact it is not used in the text.

Gauthier (1986) has presented the following solution to the prisoner's dilemma: that when the dilemma is iterated, or repeated for numerous rounds with the same people, such that the prisoners come to know the other person and his or her likely decision, then it is rational for both prisoners to cooperate, as long as each thinks that he or she is dealing with a cooperatively minded person. Otherwise, it is rational to pursue a non-cooperative strategy. Therefore, one must trust in the other's inclination to cooperate. Such trust can be achieved either by one's observations of the other person's behavior in repeated decision-making situations or by other people's testimony of the person. Gauthier has articulated that claim by distinguishing two kinds of utility-maximizing individuals: straightforward maximizers, who always seek their self-interest and refuse to cooperate, and constrained maximizers, who cooperate with fellow constrained maximizers but not with straightforward ones. Moreover, Gauthier has argued that

individuals will choose to dispose themselves as constrained maximizers instead of straightforward ones—that is, to re-train themselves not to think of their self-interest first but instead dispose themselves to honor their agreements if they find themselves in an environment of like-minded individuals (Gauthier 1986; Peterson 2009).

Applying the principle of *niṣkāmakarma* to Gauthier's solution yields a different interpretation of the prisoner's dilemma: that both prisoners should follow the principle of *niṣkāmakarma*. In Gauthier's terminology, such people are constrained maximizers because they want to maximize the common good but do not aim to maximize their individual benefits. However, that interpretation must be made with a certain reservation. One may suggest that people who follow the principle of *niṣkāmakarma* should cooperate with all persons, including straightforward maximizers or those who do not follow the principle, rather than only with like-minded people (i.e., fellow-constrained maximizers). After all, only then will the followers of *niṣkāmakarma* truly heed Kṛṣṇa's advice to not expect any selfish results. Thus, according to this interpretation, the followers of *niṣkāmakarma* should be ready to accept that they can be ruthlessly exploited by straightforward maximizers. However, in the *Bhagavadgītā*, Kṛṣṇa recommends nothing like submissiveness or passive behavior, not even when such behavior would help to avoid a war and would save people's lives. In fact, it is just the opposite (e.g., 2.37, 18.59). Therefore, and taking into account that the definitory rules of the 'game of life' are consequentialist in the *Gītā*, it can be concluded that the followers of *niṣkāmakarma* should cooperate only with like-minded people because only in that way can the common good be advanced.

Conclusion

This paper has discussed the *Bhagavadgītā*'s doctrine of *niṣkāmakarma*, or altruistic action, performed without expectation of reward. In the *Bhagavadgītā*, Kṛṣṇa proclaims that humans have rights only to actions and not to their results, whether good or bad (2.47). Therefore, humans should not desire any results whatsoever. Having distinguished interpretations of Kṛṣṇa's advice and referred to textual evidence, the paper has concluded that the idea that people have rights only to actions and not to their results should be interpreted either as the obligation to act selflessly contributing to the common good or as the wisdom to focus on acting rather than its outcomes. The prisoner's dilemma reveals that, of those two interpretations, the interpretation prioritizing the common good is more advantageous than the alternative because it maximizes the collective benefit.

Funding Information Open access funding provided by University of Vaasa (UVA).

Compliance with Ethical Standards

Conflict of Interest The author declares that he has no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the

article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aristotle. (1999). *The Nicomachean ethics*, trans. with notes by T. Irwin. Indianapolis, IN: Hackett Publishing Company.
- Brodbeck, S. (2004). Calling Kṛṣṇa's bluff: non-attached action in the *Bhagavadgītā*. *Journal of Indian Philosophy*, 32(1), 81–103.
- Chakrabarti, A. (1983). Is liberation (mokṣa) pleasant? *Philosophy East and West*, 33(2), 167–182.
- Chakraborty, S. K. (1996). Human values and Indian ethos. In W. Gasparski & L. V. Ryan (Eds.), *Human action in business: praxiological and ethical dimensions* (pp. 171–196). New Brunswick, NJ: Transaction Publishers.
- Chakraborty, S. K. (1998). Ethics for business and management: explorations in Hindu thought. In B. N. Kumar & H. Steinmann (Eds.), *Ethics in International Management* (pp. 281–308). Berlin: Walter de Gruyter.
- Chuang, C. (2015). Understanding desireless action as benevolent action. *Asian Philosophy: An International Journal of the Philosophical Traditions of the East*, 25(2), 132–147.
- Davidson, D. (1980). *Essays on actions and events*. Oxford: Oxford University Press.
- Feuerstein, G. (Trans.) (2014). *The Bhagavad-Gītā: a new translation*. Boulder, CO: Shambhala.
- Fowler, J. D. (2012). *The Bhagavad Gita: a text and commentary for students*. Eastbourne: Sussex Academy Press.
- Framarin, C. G. (2009). *Desire and motivation in Indian Philosophy*. Abingdon: Routledge.
- Gauthier, D. (1986). *Morals by agreement*. Oxford: Clarendon Press.
- Herrigel, E. (1999). *Zen in the art of archery*. New York: Vintage Books.
- Hintikka, J. (1999). *Inquiry as inquiry: a logic of scientific discovery*. Dordrecht: Springer.
- Kane, R. (1996). *The significance of free will*. Oxford: Oxford University Press.
- Kant, I. (2002). *Groundwork of the metaphysics of morals*, ed. and transl. by A. W. Wood. New Haven: Yale University Press.
- Kim, J. (1993). *Supervenience and mind: selected philosophical essays*. Cambridge, MA: Cambridge University Press.
- McKenna, M., & Pereboom, D. (2016). *Free will: a contemporary introduction*. New York: Routledge.
- Myerson, R. B. (1991). *Game theory: analysis of conflict*. Cambridge: Harvard University Press.
- Peterson, M. (2009). *An introduction to decision theory*. Cambridge: Cambridge University Press.
- Poundstone, W. (1992). *Prisoner's dilemma: John von Neumann, Game theory, and the puzzle of the bomb*. New York: Anchor Books.
- Raz, J. (1975). *Practical reason and norms*. Oxford: Oxford University Press.
- Sen, A. (2009). *The idea of justice*. Cambridge, MA: Harvard University Press.
- Singh, D. V. (1991). *Hinduism: an introduction*. Jaipur: Surabhi Prakash.
- Sreekumar, S. (2012). An analysis of consequentialism and deontology in the normative ethics of the *Bhagavadgītā*. *The Journal of Indian Philosophy*, 40(3), 277–315.
- Tilak, B. G. (1936). *Śrīmad Bhagavadgītā rahasya, or Karmayogaśāstra*, transl. by B.S. Sukthankar. Pune: Tilak and Tilak.
- van Inwagen, P. (2017). The problem of F**W*ll. In P. van Inwagen (Ed.), *Thinking about free will* (pp. 192–209). Cambridge: Cambridge University Press.
- von Wright, G. H. (1963). *Norm and action: a logical enquiry*. London: Routledge & Kegan Paul.
- Weirich, P. (2015). The lesson of the prisoner's dilemma. In M. Peterson (Ed.), *The Prisoner's Dilemma* (pp. 265–281). Cambridge: Cambridge University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.