A decision theorist's Bhagavad Gita

Harald Wiese, University of Leipzig[*]
Postfach 920, 04009 Leipzig, Germany,
tel.: 49 341 97 33 771,
fax: 49 341 97 33 779
e-mail: wiese@wifa.uni-leipzig.de

February 2016

Abstract

We analyze and interpret the *Bhagavad Gītā* from the point of view of decision theory. Árjuna asks Krishna for help in his decision of whether to fight or not. Broadly speaking, Árjuna prefers consequentialist arguments while Krishna stresses the warrior's *svadharma*. In doing so, Krishna can be considered to suggest a "new" twist on the standard decision model, in line with reason-based theories of choice. We also argue that Krishna's *svadharmic* point of view can fruitfully be seen as an example of the *Rational Shortlist Method*.

Keywords:
Bhagavad Gītā, *svadharma*, *paradharma*, reason-based choice, rational shortlist method

[*] Valuable hints have been given by participants of the research seminar of the Leipzig institute for indology and by Andre Casajus, Frank Hüttner, Katharina Lotzen, Karin Szwedek, Thomas Voss, and Ulla Wessels. I am also indebted to several anonymous referees.

Don't let the action's fruit be your motivation
Bhagavad Gītā

An act may [...] be identified with its possible consequences
Savage: The Foundations of Statistics

1 Introduction

The basic model of decision theory consists of acts, consequences and preferences. A decision maker chooses an act which leads to a consequence. Some consequences are preferred to others. As Savage (1972, p. 14) (in his major foundational work in decision theory) remarks: "If two different acts had the same consequences [...], there would [...] be no point in considering them two different acts at all." Thus, microeconomic decision theory is unabashedly consequentialist. Therefore, it may seem impossible to analyze a central writing of Hinduism, the *Bhagavad Gītā* (*Gītā* for short), from a decision-theoretic viewpoint. After all, one of Lord Krishna's famous dictums stipulates: "Don't let the action's fruits be your motivation" (*Gītā* 2.47[1]) where the Sanskrit term for fruit is *phala* which may alternatively be translated as consequence/utility/profit–terms used again and again in microeconomic texts.

The *Bhagavad Gītā* is part of book six (out of 18 books) of the great Indian epic *Mahabhárata*. The setting is this: The great warrior Árjuna is about to engage in a fight where the Pandavas (five sons of Pandu, among them Árjuna) and their allies are found on one side while the other side consists of the Kauravas (the Pandavas' cousins) together with their allies. Árjuna's charioteer is his friend Krishna who reveals himself as God Krishna later on. Árjuna realizes that many of his relatives and teachers can be found on the other side. Imagining the consequences of a deadly fight, he decides against fighting and tells Krishna about his decision. Krishna then uses many different arguments and manners to convince Árjuna that, after all, he should fight. Finally, Árjuna is convinced and the battle can begin.

Árjuna's moral dilemma can be rephrased in terms used by the famous sociologist Max Weber. He distinguishes acts that are instrumentally rational

[1] We use the relatively recent translation by Cherniak (2008) where you need to add 24 to the chapter, i.e., *Gītā* 2.47 is *Gītā* 26.47 in that book (on p. 189). See also the next paragraph.

from those that are value-rational. Arjuna's refusal to fight is built on instrumental rationality where (using Weber's (1978, pp. 26) words) "the end, the means, and the secondary results are all rationally taken into account and weighted". In contrast, Krishna (broadly) stands for value-rationality, for the "belief in the value for its own sake of some ethical, aesthetic, religious or other form of behavior, independently of its prospects of success" (see Weber 1978, pp. 25).

In this paper, we try to analyze the discussion between the two protagonists, and Krishna's preaching to Árjuna, in decision-theoretic terms. It seems to us that we might be the first to do so–the history of research on the *Gītā* given by Malinar (2007, pp. 17) does not mention any work done in this direction and we could not unearth any decision-theoretic approach on the *Gītā*, before or after 2007.

Among others, we obtain the following findings:

- A decision-theoretic reconstruction of some parts of the *Gītā* is possible. In particular, we can express the above quotation and others in a formal manner.
- It may seem that Árjuna, initially, argues in a purely consequentialist manner while Krishna argues in terms of *svadharma* (duty[2] in line with one's social standing). However, a closer reading reveals that Krishna does not shy away from consequentialist arguments.
- Krishna puts a new twist on the standard decision-theory model by pointing out that actions are not only relevant because of their consequences.
- We show how Krishna's *svadharmic* point of view can be seen as an example of the *Rational Shortlist Method* (to be introduced below).
- While Krishna's insistence on *svadharma* (duty in line with one's social standing) seems radical, less extreme versions are in use in almost all societies. We argue that Krishna's insistence on *svadharma* can be made fruitful for a new decision model that we like to call "*svadharmic* decision theory".

[2] Of course, *dharma* is a very difficult term. Olivelle (2009, pp. xlv-xlix) differentiates between six meanings of *dharma*. It seems that the *Gītā*'s discussion is concerned with Dh_4: "dharma ... belonging to or within the domain of a particular category of people or a particular goal toward which it is directed" (p. xlvii).

Our paper is related to the reason-based choice literature where people argue for specific actions, for making up their own mind or for convincing others (a survey is presented by Shafir, Simonson & Tversky 2008). Obviously, this is what happens in the *Gītā*. Krishna offers many reasons to Árjuna why the latter should indeed put up the fight. In particular, our paper can be seen as a special instance of the theory developed by Dietrich & List (2013). They present a formal model of how "motivating reasons" are "weighed" to arrive at a choice. In our context, there are two motivations. One is stressed by Árjuna and concerns the consequences for his extended family. Krishna puts his focus on the second, the *svadharma* issue. Although (or, indeed: since) our approach is very close to this reason-based theory, we find other theories (in particular the *Rational Shortlist Method*) more helpful in understanding the issues at hand. In some footnotes and also in the conclusion, we comment on the connections between the general reason-based theory put forward by Dietrich and List and our application.

The story of the *Gītā*, from the time it was composed to modern times, is presented by Davis (2015). It is well beyond the scope of this paper to discuss the *Gītā*'s relation to the *Veda*s or the *Upanishad*s, to *Sankhya* philosophy or to the *Yoga* by Patanjali. We will also not frame our discussion in terms of the often used classification of *karma yoga* (discipline of action), *jñāna yoga* (discipline of knowledge) and *bhakti yoga* (discipline of devotion). Clearly, a decision theoretical analysis is most closely related to the discipline of action. Indeed, in the section on *karma yoga*, we find Krishna's dictum that "[o]ne's own duty [*svadharma*, HW], even if done imperfectly, is better than another's [*paradharma*, HW], even if done well" (*Gītā* 3.35). However, we will also need to relate our discussion to *buddhi yoga* (discipline of understanding) and come back to this important concept in the concluding section.

The paper is organized as follows. The next section presents the basics of decision theory. We then turn to Árjuna's arguments against fighting in section 3 while Krishna's counter arguments are analyzed in section 4. In section 5, we present *svadharmic* decision theories that are less radical than the one we impute to Krishna. Section 6 concludes the paper.

2 Decision theory

2.1 Relations and preference relations

Preference relations, actions, consequences, states of the world, choice functions etc. form the ingredients of decision theory. The symbols used in this paper are listed in the appendix. We provide the necessary building blocks by borrowing freely from Kreps (1988), Rubinstein (2006), and Simon (1955, p. 102). We begin with the concept of a preference relation. It is denoted by $\succsim$ where $x \succsim y$ stands for "x is at least as good (as preferable, as virtuous, as compatible with svadharma) as y".

Definition 1 (preference relation) *Let X be any non-empty set (of "objects"). A (weak) preference relation on X is denoted by $\succsim$. Weak preference relations are said to be complete if $x \succsim y$ or $y \succsim x$ holds for all $x, y \in X$.*

The indifference relation $\sim$ is defined by

$$\begin{aligned} x &\sim y \textit{ means} \\ x &\succsim y \textit{ and } y \succsim x \end{aligned}$$

and the strict preference relation $\succ$ is defined by

$$\begin{aligned} x &\succ y \textit{ means} \\ x &\succsim y \textit{ and not } y \succsim x. \end{aligned}$$

Strict preference relations are said to be complete if $x \succ y$ or $y \succ x$ holds for all $x, y \in X$, $x \neq y$.

Completeness of preferences means that the agent "knows what he wants". Of course, in real life, this is not always the case. In this paper, we will discuss complete and incomplete preference relations. The strict preference relation $\succ$ may be complete (depending on the set X), but need not. It is not complete if there are two objects x and y in X which the agent finds equally attractive ($x \sim y$). Strict preferences obey asymmetry. This means: $x \succ y$ implies "not $y \succ x$".

It is important to note that preferences do not necessarily refer to egotistic motives. Indeed, in this paper, we are mainly concerned with moral arguments weighed by Árjuna and Krishna. Árjuna's selfish motives never play a role in his own deliberation, and only sometimes in Krishna's arguments.

2.2 Actions, consequences, and states of the world

The basic microeconomic decision model consists of

- a set of actions A,
- a set of consequences C,
- a consequence function $f : A \to C$ that attributes a consequence $c = f(a) \in C$ to an action $a \in A$, and
- a preference relation $\succsim$ on C.

In the standard decision model, an agent chooses an action $a \in A$, earns the consequence $f(a)$ which may be better or worse than consequences obtained from other actions. The theoretical prediction is an action a^* that obeys

$$\underbrace{f(a^*)}_{\in C} \succsim \underbrace{f(a)}_{\in C} \text{ for all } a \in A.$$

Differently put, the decision maker chooses an action a^* with consequence $f(a^*)$ such that no other action b exists that leads to a consequence $f(b)$ which is better than $f(a^*)$.

In this basic decision model, it does not really matter whether preferences are defined on C or on A. However, for the analysis of the *Bhagavad Gītā*, we need to distinguish between these preference definitions carefully (see subsection 2.5 below).

Sometimes, we want to consider a subset A' of the whole action set A. Let $\succ$ be an asymmetric relation on A (e.g., a strict preference relation). By

$$\max(A'; \succ) \subseteq A'$$

we then denote "best" actions from A', i.e., those actions a from A' for which no other action $b \in A'$ with $b \succ a$ exists. In particular, a^* from above is a best action from A.[3]

In more involved models, a set of states of the world W is also added. By $A \times W$, we mean the set of tuples (a, w) with $a \in A$ und $w \in W$. Instead of a consequence function f, we then deal with an uncertain-consequence

[3] Abusing notation, if $\max(A'; \succ)$ contains only one element, we will sometimes consider $\max(A'; \succ)$ an element of A', rather than a subset of A'.

function $g : A \times W \to C$, i.e., a consequence $c \in C$ is determined by both an action $a \in A$ and a state of the world $w \in W$. Often, a matrix (see fig. 1) is used to express g.[4]

		state of the world	
		state 1	state 2
decision maker	action a	$g\,(\text{action } a, \text{ state } 1)$	$g\,(\text{action } a, \text{ state } 2)$
	action b	$g\,(\text{action } b, \text{ state } 1)$	$g\,(\text{action } b, \text{ state } 2)$

Figure 1: A payoff matrix

2.3 Choice functions and WARP

Following Manzini & Mariotti (2007, p. 1826), we introduce the concept of a (point-valued) choice function.[5] The idea of a choice function is this: Consider a set of actions A and a nonempty subset $A' \subseteq A$. Now, given A', choose exactly one element from A'.

Let $\mathcal{P}(A)$ be the set of nonempty subsets of A. Formally, we have

Definition 2 (choice function) *Let A be a set of actions with $|A| > 2$. A choice function γ on A is given by*

$$\begin{aligned} \gamma &: \mathcal{P}(A) \to A, \text{ with} \\ \gamma(A') &\in A' \text{ for every } A' \in \mathcal{P}(A)\,. \end{aligned}$$

For example, if the strict preference relation $\succ$ on A is complete, $\gamma(A') = \max(A'; \succ)$ defines a choice function. Consider, however, a subset $A' = \{a, b\}$ with neither $a \succ b$ nor $b \succ a$. Then, we have $\max(A'; \succ) = A'$ and, hence, $\gamma(A') := \max(A'; \succ)$ does not define a choice function.

[4] In terms of the reason-based theory, $g\,(\text{action } a, \text{ state } 1)$ could be called an alternative as could action a (see Dietrich & List (2013, p. 106)).

[5] A set-valued definition is used by Kreps (1988, p. 12).

Choice functions γ may, or may not, obey the weak axiom of revealed preference (**WARP**): If action a is chosen in a situation where b is also feasible, then b cannot be chosen in another situation where both a and b are feasible. The idea is this: The fact that a (and not b) was chosen in the first situation tells us that a is preferred over b.

2.4 The rational short-list method

Manzini & Mariotti (2007) present and axiomatize the *Rational Shortlist Method.* According to this decision procedure, agents use two (or more) rationales in a prespecified order. Let $\succ_1$ and $\succ_2$ be asymmetric relations on A. Let A_1 be the set of actions surviving application of $\succ_1$, i.e., $A_1 = \max(A; \succ_1)$. Then, we apply $\succ_2$ to A_1 to obtain $A_2 = \max(A_1; \succ_2)$. For example, in order to choose a car, the decision maker first rejects all cars that cost more than € 10.000. Then, among the remaining cars, he chooses the one (let us assume there is only one) with the smallest milage.

Definition 3 (rational shortlist method) *A choice function γ is a rational shortlist method (RSM), if a pair of asymmetric relations $(\succ_1, \succ_2)$ exists such that*

$$\gamma(A') = \max(\max(A'; \succ_1); \succ_2)$$

holds for all $A' \in \mathcal{P}(A)$.

This definition implies that, for each subset of A, the sequential application of the two rationales leads to exactly one choice.

Manzini & Mariotti (2007) show that RSM does not, in general, obey the weak axiom of revealed preference (**WARP**). Therefore, it is somewhat lacking in rationality.

2.5 Distinguishing between four kinds of preference relations

Broadly speaking, Árjuna's arguments refer to consequences and Krishna's, to actions. Therefore, we propose to distinguish between four kinds of preferences:

1. a preference relation $\succsim_C$ on C,[6]

2. a preference relation $\succsim_A$ on A,[7] and

3. a preference relation $\succsim_{A\times C}$ on $A \times C$, where elements from $A \times C$ are action-consequence tuples $[a, c]$.[7]

Of course, actions and consequences cannot be mixed arbitrarily. Let us assume a model without states of the world (certain consequences). Then we can derive

4. another preference relation, $\succsim$ on A, by defining

$$a \succsim b :\Leftrightarrow [a, f(a)] \succsim_{A\times C} [b, f(b)] .$$

According to the fourth preference relation an action a is weakly preferred to an action b if the action-consequence tuple $[a, f(a)]$ resulting from action a is preferred to $[b, f(b)]$ by the third preference relation.

With respect to $\succsim_C$ (on C), Árjuna argues against fighting by pointing to the fierce killing involved. Krishna uses the same preference relation to convince Árjuna that shying away from fighting is bad for the latter's reputation as a fearless warrior. The preference relation $\succsim_A$ on A is used to bring home Krishna's insistance on svadharma: Irrespective of the consequences, doing one's duty is better than not doing it.

In general, preferences $\succsim_{A\times C}$ for both actions and consequences may be relevant. Finally, an action has to be chosen. A central topic of the *Gītā* is how to find and argue for preferences $\succsim$ on A (the fourth preference relation).

3 Despondent Árjuna

3.1 The Gita

Arguing for a human decision theory that deviates from more simplistic decision models, Selten (1978, pp. 147) suggests three levels of decision making:

[6] Reason-based theory deals with motives for an agent's preferences (see Dietrich & List (2013, p. 106)). Roughly, consequences (for Árjuna) and actions (for Krishna) provide these motives.

[7] Reason-based theory builds on a "regularity assumption" (see Dietrich & List (2013, p. 110)): If we have two sets of motivating reasons, then their intersection and their union also form sets of motivating reasons. Thus, with respect to the union, if consequences and actions provide motivating reasons, so do both.

the levels of (i) routine, (ii) imagination, and (iii) reasoning. Let us associate these three states with (i) the very early Árjuna, (ii) the early Árjuna, and (iii) the late Árjuna.[8]

(i) It may be argued that the very early Árjuna, willing to fight, is on the routine level. After all, fighting is a warrior's duty (*kṣatradharma*[9]).

(ii) Then, after inspecting the opposing side, the early Árjuna is horrified: "Krishna, at the sight of my own kin standing here ready to fight, my limbs feel tired and my mouth has gone dry, my body is trembling and my hair is standing on end" (*Gītā* 1.28 - 29). The warrior imagines the consequences of fighting. He expounds a cascade of consequences leading, in several steps, from (a) the destruction of the family clan (*kula*) over (b) the loss of *dharma* and (c) the cessation of offerings to ancestors to (d) eternal hell (*Gītā* 1.40 - 44). Here, Árjuna invokes *kuladharma*. The warrior is also aware that he may suffer from a bad conscience: "Better in this world to live on alms without killing the mighty elders; for were I to kill the elders, eager though they are for worldly gain, in this very world I would taste pleasures smeared with blood" (*Gītā* 2.5).

The implication drawn by the early Árjuna is clear: "It would be better for me if Dhrita·rashtra's sons [Árjuna's cousins, HW], armed with weapons, were to kill me in battle unresisting and unarmed!" (*Gītā* 1.46)

3.2 Decision-theoretic analysis

The very early Árjuna routinely considers fighting, only. One interpretation is this: Árjuna is a warrior whose action set is not

$$A = \{\text{fight, not fight}\}$$

but the smaller action set

$$A_{sv} = \{\text{fight}\}.$$

[8] Dietrich & List (2013, pp. 118-1090) discuss how propositions become motivating and mention, tentatively, three possibilities: (i) abstract conceptualization, (ii) qualitative understanding, and (iii) attention. These three points are related to our story, but we prefer a temporal account that is more in line with the *Gītā*.

[9] Both *kṣatradharma* and *kṣatriyadharma* could be used. The *Gītā* employs neither of these terms, but *kṣatradharma* shows several times in the parts of book six.

Here, *sv* refers to *svadharmic* where *dharma* is translated as duty and *sva* means "own". For the warrior Árjuna *svadharma* is *kṣatradharma* (warrior duty). In normal decision-theoretic parlance, actions from A are called feasible. Here, A is not feasible but A_{sv}, only. The restricted action set may be the result of *dharma* feasibility, rather than technical feasibility or financial feasibility in economic models.

The early Árjuna becomes aware of his full action set $A = \{\text{fight, not fight}\}$. He also contemplates on the possibility that the action "fight" might result in victory or defeat, i.e., we use action the set of states of the world

$$W = \{\text{victory, defeat}\}$$

to formalize the interplay of actions and states of the world. It seems that Árjuna does not entertain the hope that renouncing fighting might lead to his cousins' seeking an amiable solution. Therefore, the action "not fight" is automatically (with probability 1) associated with the loss of kingdom, i.e., we have

$$\begin{aligned} g\,(\text{fight, victory}) &= \text{kingdom regained and family destruction,} \\ g\,(\text{fight, defeat}) &= \text{kingdom lost and family destruction,} \\ g\,(\text{not fight}, \cdot) &= \text{kingdom lost without family destruction.} \end{aligned}$$

or the matrix of fig. 2.

		state of the world	
		victory	defeat
Arjuna	fight	kingdom regained and family destruction	kingdom lost and family destruction
	not fight	kingdom lost without family destruction	kingdom lost without family destruction

Figure 2: The early Árjuna's payoff matrix

We argue that the early Árjuna's assessment of the consequences is given by

$$\begin{aligned}&\text{kingdom regained and family destruction}\\ \prec_C\ &\text{kingdom lost without family destruction}\\ \succ_C\ &\text{kingdom lost and family destruction.}\end{aligned}$$

Árjuna says: "And we don't even know which is preferable: to vanquish or to be vanquished" (*Gītā* 2.6). "not knowing" might be translated by indifference:

$$\begin{aligned}&\text{kingdom regained and family destruction}\\ \sim_C\ &\text{kingdom lost and family destruction}\end{aligned}$$

Incomplete preferences (Árjuna is not able to rank these two consequences one way or another) provide an alternative plausible interpretation.

Be that as it may, for the early Árjuna, "not fight" is a dominant action. This means that "not fight" is better than "fight" for both states of the world.[10] Therefore, we do not need to speculate about any probabilities attached to victory or defeat, or to regaining or losing the kingdom (in case of fighting).[11]

Since the routine level (*kṣatradharma*) and the imagination level (*kuladharma*) militate for contradictory recommendations, Árjuna is despondent and does not know what to do. He turns to Krishna for help: "... my mind confused over my duty [translation of *dharma*, not *svadharma*, HW], I ask you to tell me for sure what would be best" (*Gītā* 2.7). Here, "what would be best" is clearly to be understood in terms of $\succsim$ on A.[12]

Thus, the despondent Árjuna is torn between the routine-level decision of fighting (very early Árjuna) and the imagination-level decision of not fighting (early Árjuna). One might say that Árjuna is confronted with a hard choice (see the title of a book by Levi 1986). In hard-choice situations, a decision maker sees no obvious way to come to a conclusion. Could Árjuna not just consult his preferences $\succsim$ on A (see subsection 2.5)? No, if his preferences

[10] In terms of reason-based theory, we can say that the proposition "the family is destroyed" is Árjuna's motivating reason to abstain from fighting, together with the claim that fighting is responsible.

[11] Formally, the expected utility of "not fight" would be higher than the expected utility of fight for every probability distribution on W (see, for example, Kreps 1988, pp. 31).

[12] Of course, reason-based theory fits nicely: Árjuna's asks Krishna to provide him with additional motivating reasons.

were complete (note our definition of preferences above), the deliberation process would be finished. However, this process is what the *Bhagavad Gītā* is about (from the decision-theoretic standpoint). Indeed, Kliemt (2009, pp. 48) and other philosophers of decision theory argue that complete "preferences are not reasons to act".

4 Krishna's counter-arguments

4.1 The body-as-garment argument

Turning to Selten's reasoning level, we now deal with the many different arguments Krishna uses to persuade Árjuna. While Krishna does not deny that many people might be killed (as they indeed will in great numbers), he tries to influence Árjuna's outlook on family destruction. Krishna argues that the body is of minor importance, it is the soul that counts. He uses the word *deh-in* for (embodied) soul, i.e., the (soul who is the) possessor of the body. In Krishna's words: "Whoever thinks this soul can kill or be killed, doesn't understand. It neither kills, nor is it killed. It isn't born; it never dies ... . Just as a man casts off his worn-out clothes and puts on other new ones, so the embodied soul casts off its worn-out bodies and takes other new ones" (*Gītā* 2.19 - 22).[13]

Even if Árjuna were not to accept his body-as-garment argument, Krishna has a second line of attack: "... death is certain for those who are born, and birth is certain for those who die; and so, this being inevitable, you shouldn't grieve" (*Gītā* 2.27).[14]

In terms of our theoretic model, Krishna's first arguments take exception to Árjuna's consequence function g. Firstly, souls do not die but merely take on new bodies (body-as-garment argument). Secondly, everybody dies sooner or later (death certain). Alternatively, Krishna's words are directed against Árjuna's preferences $\succsim_C$ on C with respect to the death of people, and in particular to family destruction. In any case, Krishna is telling Árjuna that the consequences of many people dying are not as serious as the latter makes them out to be. To our mind, these arguments have a consequentialist

[13] Thus, Krishna adduces the motivating reason "souls cannot be killed", together with the claim that, therefore, "fight" cannot be blamed.

[14] Krishna points to the motivating reason "people die with or without fighting" so that, again, "fight" is not responsible.

flavor.

4.2 A dominance argument

If Krishna's arguments from the first subsection are consequentialist, this is a fortiori true for the ones he then offers. Krishna points out to Árjuna the two-fold negative personal consequences of withdrawing from battle. First of all, Árjuna would miss the chance to attain heaven: "You should attend to your own duty [*svadharma*] and stand firm, for there is nothing better for a warrior than a legitimate battle. Happy the warriors who find such a battle ... –an open door to heaven ..." (*Gītā* 2.31-32). Here, it is important to note that the battle to be fought has to be a legitimate one, in line with *dharma* (both *Gītā* 2.31 and 33 stress the *dharmic* nature of the battle to come).

Second, Árjuna is warned against serious reputational damage: "The great warriors will think you withdrew from the battle out of fear, and though highly regarded by them before, you will be slighted. Your enemies too will say many unseemly things, disparaging your ability; and what could be more painful than that? Get up ... and resolve to fight! For you will either be killed and attain heaven, or you will prevail and enjoy the earth" (*Gītā* 2.35-37).[15]

With these two arguments, Krishna draws attention to consequences of Árjuna's unwillingness to fight, that might have gone unnoticed by Árjuna himself. In this manner, Krishna corrects Árjuna's view of the consequence function g. We can safely assume that the protagonists share the same preference assessment of these consequences. If we concentrate on these (as Krishna wants Árjuna to), "fight" becomes a dominant action (see the matrix of fig. 3).

4.3 Exculpation

Krishna also presents an argument against Árjuna's bad conscience ("pleasures smeared with blood"). He exculpates the hesitating warrior from the consequences of fighting by claiming: "I am Time, the world destroyer, ripened, and here I am busy crushing the worlds. Even without you, all the warriors drawn up in the opposing ranks will cease to exist. ... I have myself long since doomed them to perish; you just be the instrument ... " (*Gītā* 11.32-33).

[15] The motivational reason here is "reputation will be lost" and this loss is linked to"not fight".

		state of the world	
		victory	defeat
Arjuna	fight	prevail and enjoy the earth	be killed and attain heaven
	not fight	shameful loss of reputation	shameful loss of reputation

Figure 3: Árjuna's payoff matrix, as argued by Krishna

Thus, Krishna tells Árjuna that he is wrong about the consequences ensuing from "not fight" (see fig. 2). Árjuna cannot prevent family destruction.[16]

4.4 Equanimity

Krishna recommends equanimity to Árjuna by saying "Don't let the action's fruit be your motivation" (*Gītā* 2.47). He explains: "He whose mind is unperturbed in times of sorrow, who has lost the craving for pleasures, and who is rid of passion, fear and anger, is called a sage of steadied thought. His wisdom is secure who is free of any affections and neither rejoices nor recoils on obtaining anything good or bad" (*Gītā* 2.56-57).

To us, Krishna seems to advocate a preference relation $\succsim_C$ with

$$\text{pleasure} \sim_C \text{sorrow}.$$

Here, pleasure or sorrow do not only refer to Árjuna's egotistic motives but also to Árjuna's preferences for his *kula*. Basically, Krishna is saying that consequences are unimportant. Four comments are in order:

- One may feel that this equanimity advice stands in contrast to Krishna's warning about Árjuna's loss of reputation (see the previous subsection).
- Indifference may be too strong a notion. Sri Sankaracharya expresses in a more cautious manner: "He does not exult in pleasure, nor is he averse to pain that may befall him" (see Sastry 1977).

[16] Krishna points to the motivating reason "abstention from fighting cannot prevent family destruction".

- See our more balanced account of Krishna's perspective on consequences in the subsection on *buddhi yoga* of the concluding section.

- In the philosophical literature, there is a discussion about whether one can decide to have specific preferences or desires (see the discussion by Millgram 1998). Indeed, the economist Frank (1987) asks the question: "If homo economicus could choose his own utility function, would he want one with a conscience?" Here, we might ask the related question: "If man could choose his own utility function, would he want one governed by equanimity?" We do not pursue this discussion in this paper.

4.5 Svadharma and paradharma

While Krishna suggests equanimity with respect to $\succsim_C$, he does not do so with respect to $\succsim$ on A. "Actions, not fruits" seems to be his anti-consequentialist motto: "You have a right to the action alone, never to its fruits. Don't let the action's fruit be your motivation, and don't be attached to inactivity. ... the wise ones of disciplined understanding renounce the fruit produced by action and ... attain the perfect state" (*Gītā* 2.47-51). A little later, Krishna then specifies the action Árjuna is to perform: "One's own duty [*svadharma*, HW], even if done imperfectly, is better than another's [*paradharma*, HW], even if done well. The duty of others is fraught with danger; better to die while fulfilling one's own" (*Gītā* 3.35). Madhusudana Sarasvati (1998, pp. 252-253) explains: "The duty is one's own which is prescribed (by the scriptures) for the respective caste and stage of life."

We propose the following decision-theoretic interpretation. Let us turn to the preference relation on $A \times C$ where $[a, c] \succsim_{A\times C} [a', c']$ means that the action-consequence tuple $[a, c]$ is weakly prefered to $[a', c']$. Thus, we now admit the possibility that both actions and consequences may be relevant.

Definition 4 *Assume preferences $\succsim_{A\times C}$ on $A \times C$. They are purely consequentialist, if there is a preference relation $\succsim_C$ on C such that*

$$\begin{aligned} & [a, c] \succsim_{A\times C} [a', c'] \\ \Leftrightarrow \quad & c \succsim_C c'. \end{aligned}$$

$\succsim_{A\times C}$ are purely action-oriented if a preference relation $\succsim_A$ on A exists with

$$\begin{aligned} & [a, c] \succsim_{A\times C} [a', c'] \\ \Leftrightarrow \quad & a \succsim_A a'. \end{aligned}$$

We remind the reader of Savage's (1972, p. 14) remark: "If two different acts had the same consequences [...], there would [...] be no point in considering them two different acts at all." In a sense, the standard decision-theoretic attitude is purely consequentialist (a more balanced view is offered in the conclusions).

In contrast, Krishna's insistence on *svadharma* can be expressed by

$$\begin{aligned} A &= A_{sv} \cup A_{pa}, \\ A_{sv} \cap A_{pa} &= \emptyset \end{aligned}$$

where *pa* refers to *paradharma* or laws for others and

$$[a_{sv}, c] \succ_{A \times C} [a_{pa}, c']$$

whenever $a_{sv} \in A_{sv}$ and $a_{pa} \in A_{pa}$, for any c and $c' \in C$. Thus, Krishna's *svadharmic* point of view is a special instance of pure action orientation. It implies $[a_{sv}, f(a_{sv})] \succ_{A \times C} [a_{pa}, f(a_{pa})]$ and hence, by the fourth definition given in subsection 2.5, $a_{sv} \succ a_{pa}$.

4.6 Sequentially rationalizable choice and svadharmic RSM

We now show that Krishna's *svadharmic* point of view can be seen as an example of the *Rational Shortlist Method* (RSM) due to Manzini & Mariotti (2007) (see subsection 2.4 above). In the context of the *Bhagavad Gītā*, the *svadharma* check comes first, and the consequence check second. Thus, we consider the relation $\succ_1 = \succ_{sv}$ on A given by

$$a \succ_{sv} b :\Leftrightarrow a \in A_{sv} \wedge b \in A_{pa}$$

which is asymmetric because A_{sv} and A_{pa} are disjoint. For a subset $A' \subseteq A$, $\max(A'; \succ_{sv})$ contains only *svadharmic* actions or only *paradharmic* actions. Consider these three cases:

1. A' contains *svadharmic* and *paradharmic* actions. Then, we have $\max(A'; \succ_{sv}) = A' \backslash A_{pa}$.

2. A' contains only *svadharmic* actions. Then, we have $\max(A'; \succ_{sv}) = A'$.

3. A' contains only *paradharmic* actions. Then, we have $\max(A'; \succ_{sv}) = A'$.

Thus, the *svadharma* check is effective if and only if both *svadharmic* and *paradharmic* actions are available. This is, of course, the situation the despondent Árjuna finds himself in.

Let us assume an asymmetric relation $\succ_C$ on $f(A)$ that is complete. Consider the function $\gamma_{sv} : \mathcal{P}(A) \to A$ defined by

$$\gamma_{sv}(A') = \max(f(\max(A'; \succ_{sv})); \succ_C), A' \subseteq A.$$

It is a (well-defined) choice function by completeness of $\succ_C$. It is also an RSM because we can define $a \succ_2 b$ by $f(a) \succ_C f(b)$ and then rewrite γ_{sv} as

$$\gamma_{sv}(A') = \max(\max(A'; \succ_{sv}); \succ_2), A' \subseteq A.$$

As we have noted in subsection 2.4, RSMs do not, in general, obey **WARP**. However, we obtain the following theorem:

Theorem 1 *The choice function γ_{sv} fulfills **WARP**.*

The proof is relegated to the appendix.

5 Svadharmic decision theories

5.1 Svadharmic distance

While Krishna's *svadharmic* point of view seems extreme, less extreme versions may be interesting for decision theory in general. After all, it is held by many people that specific behaviors are, or are not, "befitting somebody's station". We briefly introduce and mention two possible subtheories, a *svadharmic* decision theory built on *svadharmic* distance (this subsection) and one with satisfycing (next subsection).

Consider a *svadharmic* distance function d on A where $d(a)$ means the distance of action a to $A_{sv} \subseteq A$. In particular, $d(a) = 0$ for every $a \in A_{sv}$. Also, very inappropriate actions a are characterized by high distances $d(a)$.

Definition 5 *Consider an action set A, a* svadharmic *subset $A_{sv} \subseteq A$ and a* svadharmic *distance function $d : A \to \mathbb{R}$ with $d(a) = 0$ for all $a \in A_{sv}$. Let $f : A \to C$ be a certain-consequence function and let $\succsim_C$ be a preference relation on $f(A) \subseteq C$. A relation $\succsim_{wsv}$ on A is called a weak* svadharmic *relation if it obeys the following properties*

- *(importance of* svadharma*): For every two actions a, b with $f(a) \sim_C f(b)$ and $d(a) < d(b)$, we have*

$$a \succ_{wsv} b$$

- *(importance of consequences): For every two actions a, b with $f(a) \succ_C f(b)$ and $d(a) = d(b)$, we have*

$$a \succ_{wsv} b$$

An agent encounters a hard svadharmic *choice between actions a and b if*

- *$d(a) > d(b)$ (b more in line with* svadharma *than a),*
- *$f(a) \succ_C f(b)$ (consequence of a preferred to consequence of b), and*
- *neither $a \succsim_{wsv} b$ nor $b \succsim_{wsv} b$*

hold.

5.2 Svadharmic satisficing

In a very influential paper, Simon (1955) argues for a satisficing decision model. Actors search for better alternatives until they happen upon an action whose consequence is deemed satisfactory. Consider the following *svadharma* version of satisficing:

Definition 6 svadharmic *satisficing is defined by the following procedure:*

Assume a minimum consequence $\hat{c}$. Keep searching until an action from the set

$$A_{sv} \cap \{a \in A : f(a) \succsim_C \hat{c}\}$$

is found. If such an action does not exist, choose any action from

$$A_{pa} \cap \{a \in A : f(a) \succsim_C \hat{c}\}$$

6 Conclusions

The topics raised by the *Bhagavad Gītā* have been attacked from quite diverging points of view: theological and philosophical (see the monograph by Malinar 2007), or psychological (see Rank 1914, Goldman 1978). This paper explores a decision-theoretical approach. Broadly speaking, one may classify the early Árjuna's point of view as consequentialist and Krishna's standpoint as action-oriented. We develop a *svadharmic* decision theory that builds on Krishna's arguments.

***Svadharmic* topics in the *Gītā* and beyond** It would have been quite possible to provide alternative citations from the *Gītā.* In particular, Krishna's teachings on *sattva, rajas*, and *tamas* (see Cherniak 2008, pp. 273) provide suitable examples. Broadly speaking, Krishna views the attitudes preferred by him as an instance of *sattva*, while he warns Árjuna against the *rajas* mode.

In this article, we focus on the *Gītā* which belongs to the sixth book of the Mahabhárata. However, Malinar (2007, pp. 35) rightly stresses that the discussion between Árjuna and Krishna is foreshadowed by somewhat similar arguments in the fifth book (see, for example Garbutt 2008). Yudhishthira's doubts and arguments focus on *kuladharma* and resemble those of the early Árjuna while Krishna himself, Kuntī (Yudhishthira's and Árjuna's mother), Vidulā (who is a woman from the *kṣatriya varna*/caste and written "Vidurā" by Malinar) and even Duryodhana (the eldest of the Pandavas' cousins) advocate the *kṣatradharma* and *svadharma* point of view.

After the war, Yudhishthira condemns the war and its consequences. Interestingly, *Cārvāka* (*Cārvāka* philosophy is often characterized as atheistic, non-Vedic, materialist, and hedonist) makes his appearance (see Heera 2011, pp. 19). He does not talk about pleasure, but seems to side the early Árjuna and the current Yudhishthira. *Cārvāka* blames Yudhishthira for the Kurukṣetra battle: "What have you gained by destroying your own people and murdering your own elders?" Finally, *Cārvāka* turns out to be a demon in disguise and burned to ashes.

***Svadharmic* decision theory as reason-based theory** One may interpret our paper as an example of the general reason-based theory proposed by Dietrich & List (2013). Under some plausible axioms presented and defended by these authors, they can show that the preferences between alternatives (for example: the actions undertaken by Árjuna) amount to preferences between combinations of motivational reasons. In the end, Árjuna is convinced by Krishna and chooses "fight". In the framework of reason-based preferences, Árjuna may have estimated (the *Gītā* does not tell which of Krishna's arguments were decisive[17]) that

- "fight" is true for the set of motivational reasons

  "the family is destroyed",
  "souls cannot be killed"

- "not fight" is true for the set of motivational reasons

  "souls cannot be killed",
  "reputation will be lost"

and Árjuna may have preferred the first set.

In subsection 3.2, we mention that complete preferences are usually not considered "reasons to act" by philosophers of decision theory. In the context of the present paper, we note that this observation holds for preferences $\succsim$ on A (or, in particular, for $\succsim_{wsv}$ on A) but not for (sub) preferences used to educe them. Preferences $\succ_{sv}$ (see subsection 4.6) or $\succsim_C$ (in subsection 5) enter the deliberation process and can be considered "reasons to act".

How new is *svadharmic* decision theory? Within reason-based theory, we have developed the concept of *svadharmic* decision theory that seems well-suited for decisions of agents in the context of status, rank, social classes and the like. However, we need to point out that standard decision theory is also capable of taking these aspects into account, albeit in a different manner. Whenever an action is considered as especially fitting or unfitting to a particular person, this fact (known to the agent and/or known to others) may be counted among the consequences of that action. Indeed, it is Krishna who alerts the fight-averse warrior Árjuna to the bad reputation that would result from a refusal to fight (see subsection 4.2).

[17] Agraval (1989, p. 139) argues that Árjuna's moral conflict is not resolved by arguments. Instead, Krishna manages to make Árjuna "look at the situation in a completely new way" by effecting "the relevant kind of radical conversion or enlightment".

The reader may also note that we did not discuss the reasons why specific acts are judged as *svadharmic*. One could argue that beneficial consequences (grosso modo or on average) provide these reasons. Then, the contrast between consequentialism and action orientation becomes less stark. When we argue for rules or *svadharma*, consequences are important. However, when an individual decision maker has to act, he should be guided by these rules without worrying about consequences. From this viewpoint, *svadharmic* decision theory and rule consequentialism (see the collection of articles in Hooker, Mason & Miller 2000) are close cousins.

Finally, *svadharmic* decision theory is obviously related to research on identity undertaken by psychologists, sociologists, and even economists. Akerlof & Kranton (2000) belongs to the third group but is clearly inspired by the other literatures.

Buddhi yoga To our mind, the most serious shortcoming of this paper is its inability to properly deal with *buddhi yoga* (discipline of understanding). We have rightly stressed Krishna's action orientation. However, Krishna did not simply (or mainly) advise Árjuna to disregard consequences. Rather, he teaches *karma yoga* together with *buddhi yoga*: "The man of disciplined understanding leaves his deeds here, both good and bad; so be disciplined in yoga. Yoga is skillfulness in action; the wise ones of disciplined understanding renounce the fruit produced by action and ... attain the perfect state" (*Gītā* 2.50-51). Here, Krishna discourages emotional attachment to results. This point is also discussed by the prominent commentator Sri Aurobindo (1995, chapter X, p. 95): "... it is because he acts ignorantly, with a wrong intelligence and therefore a wrong will ..., that man is or seems to be bound by his works; otherwise works are no bondage to the free soul." Thus, an important part of Krishna's teaching concerns the attitutes taken by acting humans. These attitudes are relevant for whether or not the actors are of disciplined understanding. A formal theory of these attitudes, so it seems to the current author, is beyond the reach of a decision-theoretic perspective. After all, the main decision-theoretic concepts are actions, consequences, and preferences on actions and/or consequences. Thus, the problem of discussing *buddhi yoga* appropriately cannot be done in the framework of this paper.

Future research We feel that our decision-theoretic interpretation of the *Gītā* focuses on some central points. However, our analysis is not complete.

For example, Krishna also argues for *svadharma* (i.e., for choosing actions from A_{sv}, only) by pointing to the simplification involved: "There is one resolute understanding here ... but the understanding of the irresolute are multifarious without limit" (*Gītā* 2.41). The interested reader may consult Rubinstein (1998, pp. 14) on simplification in the context of bounded rationality. Some readers may miss important teachings usually discussed in treatments of the *Gītā* such as Krishna's urging on Árjuna to rise above the Vedas in so far as they are constrained within the bind of the three gunas (*Gītā* 2.42-45). Similar to *buddhi yoga*, a decision-theoretic discussion seems impossible.

Future philosophical research may also try to solve Árjuna's moral dilemma. Not many scientists are bold enough to back Krishna or to back the early Árjuna. A noteworthy exception is the Indian Nobel Prize winner of 1998, the economist Amartya Sen, who published a paper in the Journal of Philosophy on "Consequential Evaluation and Practical Reason". In that paper, Sen (2000, p. 482) takes the early Árjuna's side and argues that "one must take responsibility for the consequences of one's actions and choices, and that this responsibility cannot be obliterated by any pointer to a consequence-independent duty or obligation."

7 Appendix

A List of symbols

- $x \succsim y$: (weak) preference (x is at least as good (as preferable, as virtuous, as compatible with svadharma) as y)
- $x \sim y$: indifference (x is as good as y)
- $x \succ y$: strict preference (x is better than y)
- A : set of actions
- C : set of consequences
- $A \times C$: set of action-consequence tuples $[a, c]$
- $\succsim_C$: preference on C

- $\succsim_A$: preference on A
- $\succsim_{A\times C}$: preference on $A \times C$
- $\succsim$ on A : preference on A derived from $\succsim_{A\times C}$
- $f : A \to C$: consequence function
- W : set of states of the world
- $g : A \times W \to C$: uncertain-consequence function
- $\max(A'; \succ)$: those actions from A' that do not have a better action in A'
- $\mathcal{P}(A)$: set of nonempty subsets of A
- $\gamma : \mathcal{P}(A) \to A$: choice function

B Theorem 1

For a proof of the theorem, we assume two actions a and b where a is chosen at A' while $b \in A'$ and $a \in A''$ hold. Then, it cannot be the case that $b \in A_{sv}$ and $a \in A_{pa}$ (because then a would have been eliminated in the first round at A'). Three possibilities remain:

1. $b \in A_{sv}$ and $a \in A_{sv}$. Then, both a and b survive the first round and $f(a) \succ_C f(b)$. In this case, $a \in A''$ cannot lose out against b in the second round.*

2. $b \in A_{pa}$ and $a \in A_{pa}$. Then, there is no other action c in A' with $c \in A_{sv}$ (otherwise, both a and b would have been eliminated). Therefore, we have $f(a) \succ_C f(b)$ and pursue as under 1.

3. $b \in A_{pa}$ and $a \in A_{sv}$. Then, b is eliminated in the first round under A' as well as under A'' (if b belongs to A'').

This concludes the proof.

References

Agraval, M. M. (1989). Arjuna's moral predicament, *in* B. K. Matilal (ed.), *Moral Dilemmas in the Mahabharata*, Indian Institute of Advanced Study (Rashtrapati Niwas) in association with Motilal Bnarsidass Publishers (Delhi), pp. 129–142.

Akerlof, G. A. & Kranton, R. E. (2000). Economics and identity, *The Quarterly Journal of Economics* **115**: 715–753.

Cherniak, A. (2008). *Mahabharata, Book Six, Volume One*, New York University Press and JJC Foundation.

Davis, R. H. (2015). *The Bhagavad Gita*, Princeton University Press.

Dietrich, F. & List, C. (2013). A reason-based theory of rational choice, *Nous* **47**: 104–134.

Frank, R. H. (1987). If homo economicus could choose his own utility function, would he want one with a conscience?, *American Economic Review* **77**: 593–604.

Garbutt, K. (2008). *Mahabharata, Book Five, Volume Two*, New York University Press.

Goldman, R. P. (1978). Fathers, sons and gurus: Oedipal conflict in the sanskrit epics, *Journal of Indian philosophy* **6**(4): 325–392.

Heera, B. (2011). *Uniqueness of Carvaka Philosophy in Traditional Indian Thought*, Decent Books.

Hooker, B., Mason, E. & Miller, D. E. (2000). *Morality, rules, and consequences: a critical reader*, Edinburgh University Press.

Kliemt, H. (2009). *Philosophy and Economics I: Methods and Models*, Oldenbourg Verlag MÃijnchen.

Kreps, D. M. (1988). *Notes on the Theory of Choice*, Westview Press, Boulder/London.

Levi, I. (1986). *Hard Choices*, Cambridge University Press, Cambridge (MA)/London.

Malinar, A. (2007). *The Bhagavadgita. Doctrines and Contexts*, Cambridge University Press.

Manzini, P. & Mariotti, M. (2007). Sequentially rationalizable choice, *American Economic Review* **97**: 1824–1839.

Millgram, E. (1998). Deciding to desire, *in* C. Fehige & U. Wessels (eds), *Preferences*, Walter de Gruyter, pp. 3–25.

Olivelle, P. (2009). *The Law Code of Vishnu*, Harvard University Press (Cambridge, Massachusetts et al.).

Rank, O. (1914). Myth of the birth of the hero, *The Journal of Nervous and Mental Disease* **41**(5): 110–117.

Rubinstein, A. (1998). *Modelling Bounded Rationality*, MIT Press, Princeton/Oxford.

Rubinstein, A. (2006). *Lecture Notes in Microeconomic Theory*, Princeton University Press, Princeton/Oxford.

Sarasvati, M. (1998). *Bhagavad Gita with the annotation Gudhartha-Dipika*, Advaita Ashrama. translated by Swami Gambirananda.

Sastry, A. M. (1977). *The Bhagavad Gita with the Commentary of Sri Sankaracharya*, 7 edn, Samata Books. first published 1897.

Savage, L. J. (1972). *The Foundations of Statistics*, 2 edn, Dover Publications, New York.

Selten, R. (1978). The chain store paradox, *Theory and Decision* **9**: 127–159.

Sen, A. K. (2000). Consequential evaluation and practical reason, *The Journal of Philosophy* **97**: 477–502.

Shafir, E., Simonson, I. & Tversky, A. (2008). Reason-based choice, *in* D. Kahnemann & A. Tversky (eds), *Choices, Values, and Frames*, Cambridge University Press, pp. 597–619.

Simon, H. A. (1955). A behavioral model of rational choice, *Quarterly Journal of Economics* **69**: 99–118.

Sri Aurobindo (1995). *Essays on the Gita*, Lotus Press. first published as a book in 1922 and 1928.

Weber, M. (1978). *Economy and Society*, University of California Press, Berkeley, Los Angeles, London.