A decision theorist’s Bhagavad Gita

Harald Wiese
University of Leipzig
Overview

- Introduction
- Decision theory
- Despondent Arjuna
- Krishna’s counter-arguments
- Conclusions
Introduction

- Krishna in the Bhagavad Gita:
  Don’t let the actions’s fruit be your motivation

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

Nevertheless: A decision-theoretic interpretation of the Gita is possible.
Bhagavad Gita (Gita for short) = part of book 6 (out of 18 books) of the Mahabhárata.

The setting is this:

- Árjuna and his brothers are about to fight against their cousins.
- 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.
- He tells Krishna that he does not want to fight.
- Krishna argues that the warrior Árjuna should fight.
- Finally, Árjuna is convinced and the battle can begin.
Definition (preference relation)

\( X \) is a non-empty set (of “objects”).
\( \succeq \) is a (weak) preference relation on \( X \)
\( x \succeq y \) means “\( x \) is at least as good (as preferable, as virtuous) as \( y \)”.
\( \succeq \) complete if \( x \succeq y \) or \( y \succeq x \) holds for all \( x, y \in X \).

Completeness = the agent “knows what he wants”

Also:

- strict preference, expressed by \( x \succ y \)
- indifference, expressed by \( x \sim y \)
The basic microeconomic decision model consists of

- a set of actions $A$,
- a set of consequences $C$,
- a consequence function $f : A \rightarrow C$ and
- a preference relation $\prec$ on $C$.

An agent chooses

- an action $a \in A$,
- earns the consequence $f(a)$.

The theoretical prediction is an action $a^*$ that obeys

$$f(a^*) \succeq f(a) \quad \text{for all} \quad a \in A.$$
Often, a set of states of the world $W$ is added.

- **Not**; consequence function $f : A \rightarrow C$
- **But**; uncertain-consequence function $g : A \times W \rightarrow C$

### state of the world

<table>
<thead>
<tr>
<th></th>
<th>state 1</th>
<th>state 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>decision maker</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>action $a$</td>
<td>$g$ (action $a$, state 1)</td>
<td>$g$ (act. $a$, st. 2)</td>
</tr>
<tr>
<td>action $b$</td>
<td>$g$ (action $b$, state 1)</td>
<td>$g$ (act. $b$, st. 2)</td>
</tr>
</tbody>
</table>
Four kinds of preferences:

- a preference relation \( \succeq_C \) on \( C \)
- a preference relation \( \succeq_A \) on \( A \) and
- a preference relation \( \succeq_{A \times C} \) on \( A \times C \).

Since actions and consequences cannot be mixed arbitrarily, we can derive

- a fourth preference relation, \( \succeq \) on \( A \), by defining

\[ a \succeq b \text{ as } [a, f(a)] \succeq_{A \times C} [b, f(b)] \]

Gita:

- \( \succeq_C \) on \( C \): Árjuna’s argument against fighting and killing
- \( \succeq_A \) on \( A \): Krishna’s insistence on svadharma
- \( \succeq_{A \times C} \) on \( A \times C \) : Both actions and consequences relevant
- \( \succeq \) on \( A \): Which action is best?
Reinhard Selten suggests three levels of decision making:

(i) routine,
(ii) imagination, and
(iii) reasoning.
Despondent Árjuna

(i) Árjuna (i) is willing to fight

Decision-theoretic interpretation:
not $A = \{\text{fight, not fight}\}$, but $A_{sv} = \{\text{fight}\}$
$s v = svadharmic$ where $d h a r m a = \text{duty/religion}$ and $s v a = \text{own}$
(ii) After inspecting the opposing side, the early Áṛjuna 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.”
- “It would be better for me if Dhrita-rashtra’s sons, armed with weapons, were to kill me in battle unresisting and unarmed!”
Despondent Árjuna
imagination level (interpretation)

- Full action set $A = \{\text{fight, not fight}\}$.
- Set of states of the world $W = \{\text{good luck, bad luck}\}$.
- Uncertain-consequence function $g : A \times W \rightarrow C$, defined by

\[
\begin{array}{c|c|c}
\text{state of the world} & \text{good luck} & \text{bad luck} \\
\hline
\text{fighting} & \text{victory and family destr.} & \text{defeat and family destr.} \\
\text{not fighting} & \text{defeat without family destr.} & \text{defeat without family destr.} \\
\end{array}
\]

For Árjuna (ii), not fighting is a dominant action.
Despondent Árjuna
incomplete preferences

- The routine level (kṣatradharma) and
- the imagination level (kuladharma)

militate for contradictory recommendations. Árjuna turns to Krishna for help:
“... my mind confused over my duty [dharma, not svadharma],
I ask you to tell me for sure what would be best.”

**Decision-theoretic interpretation:**
“what would be best” is clearly to be understood in terms of $\succeq$ on $A$.
Could Árjuna not just consult his preferences $\succeq$ on $A$?
No, they are incomplete—he does not know what to prefer.
Krishna’s counter-arguments

The body-as-garment argument

Kishna:
“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 [*deh-in*] casts off its worn-out bodies and takes other new ones.”

Decision-theoretic interpretation:
Krishna argues against Árjuna’s preferences $\succeq_C$ on $C$. 
Krishna’s counter-arguments
A dominance argument I

1. 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, Partha—an open door to heaven ...”

2. Danger of suffering 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, son of Kunti [Kunti is Árjuna’s mother], and resolve to fight! For you will either be killed and attain heaven, or you will prevail and enjoy the earth.”
**Decision-theoretic interpretation:**

Krishna corrects Árjuna’s view of the uncertain-cons. function $g$:

<table>
<thead>
<tr>
<th>State of the World</th>
<th>Good Luck</th>
<th>Bad Luck</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fighting</td>
<td>Prevail and enjoy the earth</td>
<td>Be killed and attain heaven</td>
</tr>
<tr>
<td>Not Fighting</td>
<td>Shameful loss of reputation</td>
<td>Shameful loss of reputation</td>
</tr>
</tbody>
</table>

For Krishna, fighting is a dominant action.
Krishna’s counter-arguments
Exculpation

Árjuna is afraid of his bad conscience:
“Better in this world to live an 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.”

Krishna exculpates him:
“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 ... ”.

Decision-theoretic interpretation:
Krishna tells Árjuna that he is wrong about the consequences. Árjuna cannot prevent family destruction.
Krishna’s counter-arguments

Equanimity

Krishna:

- “Don’t let the actions’s fruit be your motivation”
- “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.”

**Decision-theoretic interpretation:**
Krishna seems to advocate a preference relation $\succeq_C$ 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*. Problem: Can people choose preferences?
Krishna’s counter-arguments

Svadharma and paradigm 1

Krishna suggests equanimity with respect to $\sim C$, but certainly not with respect to $\sim$ on $A$.

“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.”

“One’s own duty [svadharma], even if done imperfectly, is better than another’s [paradharma], even if done well. The duty of others is fraught with danger; better to die while fulfilling one’s own.”
Decision-theoretic interpretation:

**Definition**

Preferences $\succeq_{A \times C}$ on $A \times C$ are purely consequentialist, if there is a preference relation $\succeq_C$ on $C$ such that

$$[a, c] \succeq_{A \times C} [a', c'] \iff c \succeq_C c'.$$

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

$$[a, c] \succeq_{A \times C} [a', c'] \iff a \succeq_A a'.$$
Krishna’s counter-arguments
Svadharma and paradharma III

Krishna’s insistence on *svadharma* can be expressed by

\[ A = A_{sv} \cup A_{pa}, \]
\[ A_{sv} \cap A_{pa} = \emptyset \]

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 \( c' \in C \).
RSM = rational shortlist method

- RSM in general: Agents use two rationales in a prespecified order. Example: Buy a car by
  1. reject all cars above € 10,000 (if one below exists) and,
  2. choose the one with the smallest milage
(Manzini and Mariotti, AER 2007).
- Svadharmic RSM:
  1. discard all paradharmic actions (if a svadharmic one exists).
  2. choose the one with the best consequences.

**Theorem**

The svadharmic RSM always fulfills the weak axiom of revealed preference.

Details in paper.
Conclusions I

- A decision-theoretic reconstruction of some parts of the Bhagavad Gita possible.
- Krishna does not shy away from consequentialists arguments.
- Krishna: actions are not only relevant because of their consequences.
- Krishna’s svadharmic point of view an example of the *Rational Shortlist Method*.
- Milder forms of svadharma (duty in line with one’s social standing, behavior “befitting one’s station”) are in use in all societies.
- Svadharmaic decision theory closely related to research on identity undertaken by psychologists, sociologists, and even economists.
Standard decision theory could also take *svadharma* aspects into account. If an action is 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.

*Svadharmic* decision theory and rule-based consequentialism may be close cousins:

- When we argue for rules or *svadharma*, consequences are important.
- Whenever an individual decision maker has to act, he should be guided by these rules, not by consequences.
Alternative citations from the Gita might have been chosen. See Krishna’s teachings on sattva, rajas, and tamas. Krishna warns Árjuna against the rajas mode.

Similar arguments in the fifth book. Yudhishthira’s doubts and arguments focus on kuladharma while Krishna himself, Kuntī and even Duryodhana advocate the kṣatradharma and svadharma point of view.

Sen’s advocacy of the early Árjuna’s position
The Indian Nobel price winner of 1998, Amartya Sen takes Árjuna’s side: “one must take responsibility for the consequences of one’s actions and choices, and [...] this responsibility cannot be obliterated by any pointer to a consequence-independent duty or obligation.”
Choice function, axiom

**Definition (choice function)**

Let \( A \) be a set of actions with \(|A| > 2\). A choice function \( \gamma \) on \( A \) is given by

\[
\gamma : \mathcal{P}(A) \to A, \text{ with } \\
\gamma (A') \in A' \text{ for every } A' \in \mathcal{P}(A).
\]

**WARP (weak axiom of revealed preference):**

If

- two actions \( a \) and \( b \) with \( a, b \in A' \)
- \( a \in A \) is chosen for \( A' \) (\( a = \gamma (A') \))
- \( a \in A'' \)

then \( b \neq \gamma (A'') \)
WARP can be violated by RSM

Three routes a, b, and c

- Traffic:
  - c less traffic than a
  - other pairs are difficult to compare

- Length:
  a shorter than b shorter than c

If

- two actions a and b with $a, b \in A' = \{a, b\}$
- $a \in A$ is chosen for $A'$ ($a = \gamma(A')$)
- $a \in A'' = \{a, b, c\}$

then $b = \gamma(A'')$, in violation of WARP
The svadharmic RSM always fulfills the weak axiom of revealed preference.

Assume
- two actions \( a \) and \( b \) with \( a, b \in A' \)
- \( a \) is chosen at \( A' \)
- \( a \in A'' \)

Then, exclude \( b \in A_{sv} \) and \( a \in A_{pa} \) (why?)

1. \( b \in A_{sv} \) and \( a \in A_{sv} \). Then, both \( a \) and \( b \) survive the first round and \( f(a) \succ_C f(b) \). Hence \( b \neq \gamma(A'') \).

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, \( f(a) \succ_C f(b) \) and see 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.