Soft Moral Choice

Erich Rast
erich@snafu.de

IFL, Universidade Nova de Lisboa

December 12, 2012

Overview

Goals

- Moral Theory and Decision Theory

Goals

- Moral Theory and Decision Theory
  - How to make rational *moral* decisions?

Goals

- Moral Theory and Decision Theory
    - How to make rational *moral* decisions?
    - In other words: How to combine moral theories with rational decision making?

Goals

- Moral Theory and Decision Theory
  - How to make rational *moral* decisions?
  - In other words: How to combine moral theories with rational decision making?
- Soft Moral Constraints

Goals

- Moral Theory and Decision Theory
  - How to make rational *moral* decisions?
  - In other words: How to combine moral theories with rational decision making?
- Soft Moral Constraints
  - How can a certain amount of permeability be allowed?

Goals

- Moral Theory and Decision Theory
  - How to make rational *moral* decisions?
  - In other words: How to combine moral theories with rational decision making?
- Soft Moral Constraints
  - How can a certain amount of permeability be allowed?
  - In other words: Can decision theory give recommendations for making rational moral decisions without describing a moral pedant?

Disclaimer

- There are many perceivable ways to soften *moral rules*.
- In the context of multi-criteria decision making it is more natural to think of the moral rules as pedantic and providing *moral values* who are dealt with in a permeable way.
- *Thresholds* will be used for that purpose.
- The values of thresholds and how to obtain them is left open in this talk and the article.
- My overall stance is meta-ethical.

What is a Value?

- Value $\approx$ worth of something ($\neq$ monetary value)
- No agreed definition can be found in the literature.
- Some important Literature (loosely grouped):
  - Brentano School: Ehrenfels (1897), Meinong (1894)
  - Scheler (1913), Perry (1954), Harman (2000)
  - American Pragmatism: Dewey (1939), Lewis (1946)
  - Scandinavian Tradition: von Wright (1963), Hansson (2001), Rabinowicz (2008)

Von Wright's Taxonomy

Lit. Von Wright, George Henrik: The Varieties of Goodness. 1963. (based on lectures given 1958-60)

- Instrumental Goodness; e.g. a knife is good for cutting meat
- Technical Goodness; e.g. a good doctor (good at healing patients, among other factors)
- The Beneficial and Medical Goodness; e.g. a certain amount of water is beneficial for a flower
- Utilitarian Goodness: the useful, advantageous, and favorable; e.g. solid math skills are good for an engineer, exercise is good for a runner
- Hedonic Goodness: pleasure, happiness, etc.; e.g. a good wine, a great holiday

Von Wright's Taxonomy

Lit. Von Wright, George Henrik: The Varieties of Goodness. 1963. (based on lectures given 1958-60)

- Instrumental Goodness; e.g. a knife is good for cutting meat
- Technical Goodness; e.g. a good doctor (good at healing patients, among other factors)
- The Beneficial and Medical Goodness; e.g. a certain amount of water is beneficial for a flower
- Utilitarian Goodness: the useful, advantageous, and favorable; e.g. solid math skills are good for an engineer, exercise is good for a runner
- Hedonic Goodness: pleasure, happiness, etc.; e.g. a good wine, a great holiday

Social norms in von Wright (1963) are complex and 'quasi-contractual.' (my words)

Jackendoff's Taxonomy

Lit. Jackendoff, Ray: Language, Consciousness, Culture. MIT Press 2007. *also*: The Peculiar Logic of Value. Journal of Cognition and Culture, Vol. 6 (2006), 375-407. Part of table taken from book p. 301.

Type of Value	Domain
Affective Value	events, situations
Utility	events, situations
Resource Value	objects
Quality	events, objects
Prowess	persons
Normative Value	action+person
Personal Normative Value	persons
Esteem	persons

Multi-criteria Decision Theory

Lit. Eisenführ, Franz & Weber, Martin & Langer, Thomas: Rational Decision Making. Springer 2010. There are many other good introductions.

- Important works: Ramsey (1931), de Finetti (1937, 1974), von Neuman & Morgenstern (1947), Savage (1954), Debreu (1959), Fishburn (1970), Keeney & Raiffa (1976)
- Types of Decision Theory (DT)
  - Normative DT: Advises how an agent should make rational choices.
  - Prescriptive DT: Recommends the best alternative given data on the decision makers preferences, risks and uncertainties involved, focusing on elicitation methods.
  - Descriptive DT: Describes how rational decision makers come up with a decision.

Why Decision Theory?

- Normative Aspect ($\neq$ moral normativity)
  - From a normative point of view DT describes the optimal choice of action on the basis of given preferences. If you make another choice, you make a mistake in the sense of violating your own preferences.
  - As Savage (1954) has shown, probabilities can be derived from qualitative preferences and purely subjective credences, so they are not needed as a basic notion.
- Avoiding Biases
  - Extensive work in psychology and behavioral economics has shown that descriptive DT is too idealizing.
  - Human decision making involves mistakes and is based on cognitive biases.
  - Some biases might be useful heuristics—work most of time—, yet others are clearly reasoning mistakes.
  - Effects of biases can be reduced by using sound elicitation methods, consistency checks, sensitivity tests, and other instruments of prescriptive DT.

Biases: The Asian Disease Example

The following example of *framing* is taken from Eisenfür (2010), p. 8 and based on Tversky & Kahneman (1981).

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people.

Biases: The Asian Disease Example

The following example of *framing* is taken from Eisenfür (2010), p. 8 and based on Tversky & Kahneman (1981).

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people.There are two alternative programs.

1. If program A is adopted, 200 people will be saved. If program B is adopted, there is a one-third probability that all 600 people will be saved and a two-thirds probability that no one will be saved. Which do you prefer, Program A or Program B?

Biases: The Asian Disease Example

The following example of *framing* is taken from Eisenfür (2010), p. 8 and based on Tversky & Kahneman (1981).

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people.There are two alternative programs.

1. If program A is adopted, 200 people will be saved. If program B is adopted, there is a one-third probability that all 600 people will be saved and a two-thirds probability that no one will be saved. Which do you prefer, Program A or Program B?

⇒ Most people choose program A.

Biases: The Asian Disease Example

The following example of *framing* is taken from Eisenfür (2010), p. 8 and based on Tversky & Kahneman (1981).

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people.There are two alternative programs.

2. If Program A is adopted, 400 people will die. If Program B is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 will die. Which do you prefer, Program A or Program B?

Biases: The Asian Disease Example

The following example of *framing* is taken from Eisenfür (2010), p. 8 and based on Tversky & Kahneman (1981).

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people.There are two alternative programs.

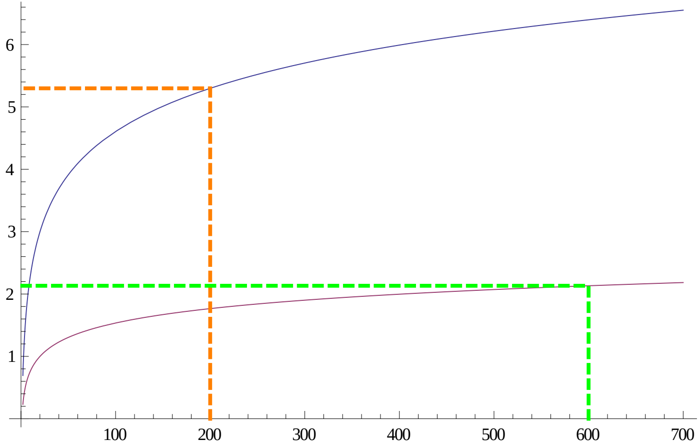
2. If Program A is adopted, 400 people will die. If Program B is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 will die. Which do you prefer, Program A or Program B?

$\Rightarrow$ Most people choose program B.

What Does Decision Theory Say?

- Situation 1 and Situation 2 are equivalent.
- In both situations the expected values of the alternatives A and B are the same: EB(A)=EV(B)=200.
- Recommendation depends on risk attitude:
  - Concave Utility Function: risk averse
  - Convex Utility Function: risk prone
  - Linear Utility Function: risk neutral
- A reasonable decision maker will, under most circumstances, be slightly risk averse.
- Notice: Money-pump arguments can exclude risk prone decisions, but do not speak for risk aversity.
- In other words, standard DT does not provide any rational justification for choosing A over B.

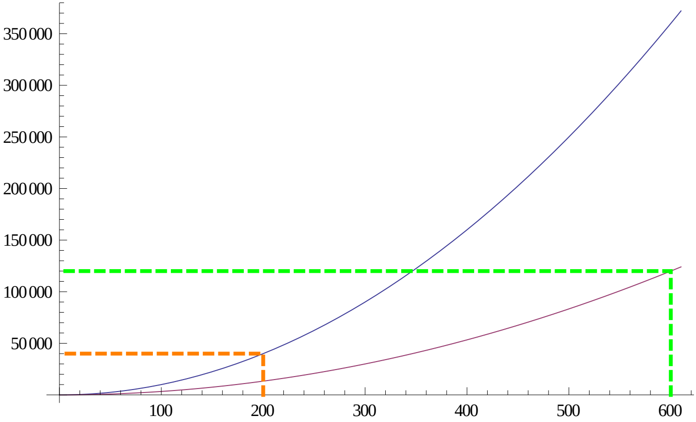
Risk Aversion



Logarithmic utility function:

$$\ln 200 \text{ (orange)} > \frac{1}{3} \cdot \ln 600 \text{ (green)}$$

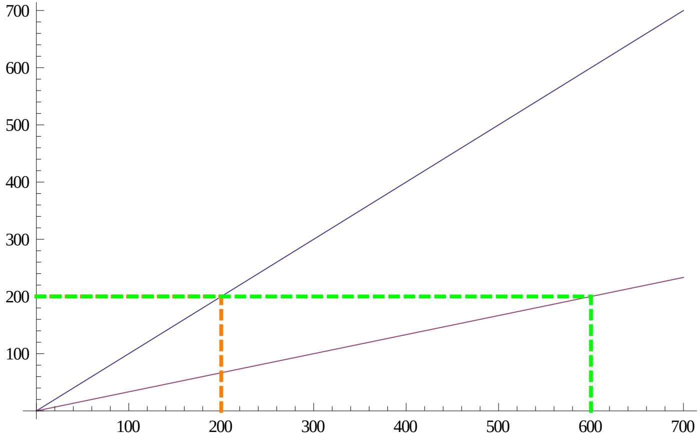
Risk Proneness



Quadratic utility function:

$$200^2 \text{ (orange)} < \frac{1}{3} \cdot 600^2 \text{ (green)}$$

Risk Neutrality



Linear utility function:

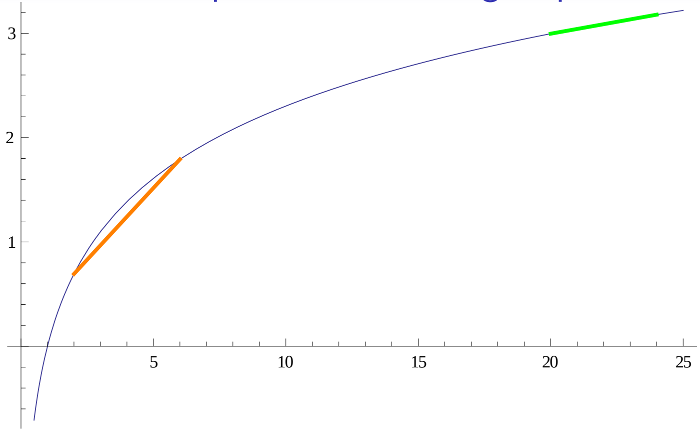
$$200 \text{ (orange)} = \frac{1}{3} \cdot 600 \text{ (green)}$$

Conceptual Problem: the Dual Role of Utility

- According to the *Principle of Diminishing Marginal Utility* many real-world a value functions are concave. The more you get from something, the less subjective worth will it have in relation to other goods.
- For example: If you have 2 bananas, getting 4 more bananas will be worth more than getting 4 more bananas if you already have 20 of them.
- The dual role of utility has been criticized from time to time; the issue is complicated.
- Common utility elicitation methods use choices between lotteries. They usually cannot distinguish between the two roles (nor can all decision makers).

Lit. The Principle of Diminishing Marginal Utility was first formulated by Gossen (1854) under a different name as a basic characteristics of human pleasure (Genuss) in the context of a pleasure-utilitarian foundations of economics.

Example: Value of Burgers per Week



Logarithmic value function:

$$\ln 6 - \ln 2 > \ln 24 - \ln 20$$

Shifting from 2 to 6 burgers is more desirable to that person than shifting from 20 to 24 burgers.

Multi-Criteria Decision Making: Additive Models

Ingredients:

- a finite set of alternatives $a, b, c, \ldots$

Multi-Criteria Decision Making: Additive Models

Ingredients:

- a finite set of alternatives $a, b, c, \ldots$
- a finite number of n attributes for each alternative, usually indexed $a_1, \ldots, a_n$ for alternative a, $b_1, \ldots, b_n$ for alternative b, ...

Multi-Criteria Decision Making: Additive Models

Ingredients:

- a finite set of alternatives $a, b, c, \ldots$
- a finite number of n attributes for each alternative, usually indexed $a_1, \ldots, a_n$ for alternative a, $b_1, \ldots, b_n$ for alternative b, ...
- weights $w_1, \ldots, w_n$ for the attributes

Multi-Criteria Decision Making: Additive Models

Ingredients:

- a finite set of alternatives $a, b, c, \ldots$
- a finite number of n attributes for each alternative, usually indexed $a_1, \ldots, a_n$ for alternative a, $b_1, \ldots, b_n$ for alternative b, ...
- weights $w_1, \ldots, w_n$ for the attributes
- for each attribute a function into $\mathbb{R}$:

Multi-Criteria Decision Making: Additive Models

Ingredients:

- a finite set of alternatives $a, b, c, \ldots$
- a finite number of n attributes for each alternative, usually indexed $a_1, \ldots, a_n$ for alternative a, $b_1, \ldots, b_n$ for alternative b, ...
- weights $w_1, \ldots, w_n$ for the attributes
- for each attribute a function into $\mathbb{R}$:
  - $v_i(.)$ called *value function* for decision making under certainty

Multi-Criteria Decision Making: Additive Models

Ingredients:

- a finite set of alternatives $a, b, c, \ldots$
- a finite number of n attributes for each alternative, usually indexed $a_1, \ldots, a_n$ for alternative a, $b_1, \ldots, b_n$ for alternative b, ...
- weights $w_1, \ldots, w_n$ for the attributes
- for each attribute a function into $\mathbb{R}$:
  - $v_i(.)$ called *value function* for decision making under certainty
  - $u_i(.)$ called *utility function* for decision making under risk

Multi-Criteria Decision Making: Additive Models

Ingredients:

- a finite set of alternatives $a, b, c, \ldots$
- a finite number of n attributes for each alternative, usually indexed $a_1, \ldots, a_n$ for alternative a, $b_1, \ldots, b_n$ for alternative b, ...
- weights $w_1, \ldots, w_n$ for the attributes
- for each attribute a function into $\mathbb{R}$:
  - $v_i(.)$ called *value function* for decision making under certainty
  - $u_i(.)$ called *utility function* for decision making under risk
- for risk: probabilities for sequences of attributes of each alternative a, jointly characterizing a probability measure over the alternatives

Multi-Criteria Decision Making: Additive Models

Ingredients:

- a finite set of alternatives $a, b, c, \ldots$
- a finite number of n attributes for each alternative, usually indexed $a_1, \ldots, a_n$ for alternative a, $b_1, \ldots, b_n$ for alternative b, ...
- weights $w_1, \ldots, w_n$ for the attributes
- for each attribute a function into $\mathbb{R}$:
  - $v_i(.)$ called *value function* for decision making under certainty
  - $u_i(.)$ called *utility function* for decision making under risk
- for risk: probabilities for sequences of attributes of each alternative a, jointly characterizing a probability measure over the alternatives

Note: The last point is a shortcut. In Savage (1954) events, actions, and their consequences are modeled explicitly.

For the present purposes, the difference does not matter.

Making a Decision: Additive Models

Decision Under Certainty

1. Compute the weighted sum of each alternative.

$$v(a) = \sum_{i=1}^{n} w_i v_i(a_i)$$

2. Choose the alternative whose value is highest.

Decision Under Risk

1. Compute the weighted sum of each alternative for sequences of attributes of the alternative and their corresponding probability.

$$u(a) = \sum_{j=1}^{m} p_j \sum_{i=1}^{n} w_{i,j} u_{i,j}(a_{i,j})$$

2. Choose the alternative whose utility is highest.

Example: Looking for a Job

Adopted from Eisenführ (2010), Fishburn (1991). Certainty is assumed in what follows.

Alternatives and Attributes

	a	b	c
1 - salary	\$65 000	\$50 000	\$58 000
2 - reputation	Low	High	Mediocre

Normalized Table with Weighted Sums

Weights: $w_1 = \frac{3}{4}$, $w_2 = \frac{1}{4}$; values for 2: $v(low) = 0.0$, $v(high) = 1.0$, $v(mediocre) = 0.4$

Sample calculations: $0.75 \cdot 0.769 + 1.0 \cdot 0.25 \approx 0.827$; $\frac{1}{65000} \cdot 58000 \approx 0.892$

	a	b	c
1	1.000	0.769	0.892
2	0.000	1.000	0.400
Weighted Sum:	0.750	0.827	0.769

Caveat 1: Risk $\neq$ Uncertainty

Risk

A risk is given when the probability of an event is known or can be estimated to a precision that is sufficient for the decision task.

Uncertainty

Uncertainty may mean many things: The probability of an event is not known at all (complete ignorance), only interval-estimated, based on uncertain inferences, based on disagreeing judgments of experts in group decision making, ...

- The simple model *cannot* deal with uncertainty, yet in real-life situations some form of uncertainty is almost always present and risks are almost never known precisely.
- Various ways to deal with uncertainty have been developed, most of which are interval-based.

Caveat 2: High Requirements

Paraphrased from Eisenführ et. el. (2010), p. 130.

$$a = \langle a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n \rangle \quad (1)$$
$$b = \langle a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n \rangle \quad (2)$$
$$a' = \langle a'_1, \ldots, a'_{i-1}, a_i, a'_{i+1}, \ldots, a'_n \rangle \quad (3)$$
$$b' = \langle a'_1, \ldots, a'_{i-1}, b_i, a'_{i+1}, \ldots, a'_n \rangle \quad (4)$$

Preferential Independence

$a \succ b \Leftrightarrow a' \succ b'$

Difference Independence

$a \rightarrow b \sim a' \rightarrow b'$

where $a \rightarrow b$ informally means 'switching from a to b.' (details skipped)

Moral Values in Decision Theory?

- Suppose there are two kinds of attributes, the moral attributes M and normal attributes N. Moral attributes are those that are *morally relevant* in a given decision situation.
- Different Points of Views
  - Deontic view: An attribute is morally relevant if it concerns an aspect of a possible course of action that is governed by moral rules.
  - Virtue-ethical view: An attribute is morally relevant if it concerns a moral aspect of the decision makers attitudes.
  - Utilitarian view: Every attribute is morally relevant?
- The formal moral value, i.e. the outcomes of $v(.)$ or $u(.)$, expresses the moral aspects of choices.

From Goodness to Values (Valuations)

The value taxonomies are based on implicit comparisons / preference relations (cf. Hansson 2001). Example for von Wright (1963):

- instrumental: comparing instruments w.r.t purposes, e.g. knifes $\Rightarrow$ consumer theory
- technical: technical: comparing abilities and professions, e.g. doctors $\Rightarrow$ consumer theory, general decision theory
- beneficial & medical: comparing factors needed w.r.t. certain goals $\Rightarrow$ general decision theory
- hedonic: comparing aspects of personal tastes $\Rightarrow$ general decision theory

From Goodness to Values (Valuations)

The value taxonomies are based on implicit comparisons / preference relations (cf. Hansson 2001). Example for von Wright (1963):

- instrumental: comparing instruments w.r.t purposes, e.g. knifes $\Rightarrow$ consumer theory
- technical: technical: comparing abilities and professions, e.g. doctors $\Rightarrow$ consumer theory, general decision theory
- beneficial & medical: comparing factors needed w.r.t. certain goals $\Rightarrow$ general decision theory
- hedonic: comparing aspects of personal tastes $\Rightarrow$ general decision theory

Additional note: During problem analysis, prescriptive decision theory focuses on *ends* of actions rather than *means*, but it can be used for both.

Moral Choice

① Moral Choice

What is the relationship between moral and non-moral attributes?

② Soft Moral Choice

How can one make rational moral choices without being a moral pedant?

Answers given (assumptions):

① The rationality requirements of moral choices are the same as those for choices in general.

② Thresholds can be used to make moral choices permeable.

What's the Problem?

Why should moral choices be 'permeable' and what does this mean anyway? Shouldn't a *moral* decision maker simply be characterized as the one who makes *moral* decisions?

What's the Problem?

Why should moral choices be 'permeable' and what does this mean anyway? Shouldn't a *moral* decision maker simply be characterized as the one who makes *moral* decisions?

Problems:

- Unless moral rules are already 'soft' themselves, nobody makes perfectly moral decisions *all the time*.
- Moral permeability cannot be characterized as 'acting morally most of the time':
  - A murder cannot be excused by the fact that the murderer acts morally most of the time.
  - Probably most murderers do so.

What's the Problem?

Why should moral choices be 'permeable' and what does this mean anyway? Shouldn't a *moral* decision maker simply be characterized as the one who makes *moral* decisions?
Problems:

- Unless moral rules are already 'soft' themselves, nobody makes perfectly moral decisions *all the time*.
- Moral permeability cannot be characterized as 'acting morally most of the time':
  - A murder cannot be excused by the fact that the murderer acts morally most of the time.
  - Probably most murderers do so.

⇒ Idea: Instead of relaxing the rules ('stealing is usually bad'), we keep them fixed ('stealing is bad') and introduce thresholds for the utility/value functions instead.

Wait a minute. . .

Some motivation...

Strict Moral Choice and Attribute Dominance

M: set of indices of moral attributes
A decision matrix is strictly moral iff. for all alternatives a, b:

$$\text{If } \sum_{i \in M} w_i v_i(a_i) > \sum_{i \in M} w_i v_i(b_i), \text{ then } v(a) > v(b) \tag{5}$$

Strict Moral Choice and Attribute Dominance

M: set of indices of moral attributes
A decision matrix is strictly moral iff. for all alternatives a, b:

$$\text{If } \sum_{i \in M} w_i v_i(a_i) > \sum_{i \in M} w_i v_i(b_i), \text{ then } v(a) > v(b) \tag{5}$$

Problem: Non-moral attributes only play a role in this model if the sums of values of all moral attributes of two alternatives a, b are exactly equal. If not, they are dispensable. Moral attributes dominate non-moral attributes.

Global Threshold View (Unipolar)

A decision matrix is moral with global threshold α iff. for all alternatives a, b:

$$\text{If } \sum_{i \in M} w_i v_i(a_i) > \alpha \text{ and } \sum_{i \in M} w_i v_i(a_i) > \sum_{i \in M} w_i v_i(b_i), \qquad (6)$$

$$\text{then } v(a) > v(b)$$

Global Threshold View (Unipolar)

A decision matrix is moral with global threshold α iff. for all alternatives a, b:

$$\text{If } \sum_{i \in M} w_i v_i(a_i) > \alpha \text{ and } \sum_{i \in M} w_i v_i(a_i) > \sum_{i \in M} w_i v_i(b_i), \tag{6}$$
$$\text{then } v(a) > v(b)$$

Problem: It seems hard in practice to elicit a global threshold for joint moral attributes from a 'moral expert' or derive it from some moral theory.

Individual Thresholds View (Unipolar)

α_i: thresholds for all $i \in M$
A decision matrix is moral with individual thresholds α_i iff. for all alternatives a, b:

$$\text{If } \exists\alpha_i \text{ s.t. } w_i v_i(a_i) > \alpha_i \text{ and } \sum_{i \in M} w_i v_i(a_i) > \sum_{i \in M} w_i v_i(b_i), \tag{7}$$

then $v(a) > v(b)$

Individual Thresholds View (Unipolar)

α_i: thresholds for all $i \in M$
A decision matrix is moral with individual thresholds α_i iff. for all alternatives a, b:

$$\text{If } \exists \alpha_i \text{ s.t. } w_i v_i(a_i) > \alpha_i \text{ and } \sum_{i \in M} w_i v_i(a_i) > \sum_{i \in M} w_i v_i(b_i), \quad (7)$$

then $v(a) > v(b)$

Reflects the way moral aspects of decisions are often evaluated in court, e.g. taking into account whether a crime has been committed out of a desperate need or sheer selfishness.

Individual Thresholds View (Unipolar)

α_i: thresholds for all $i \in M$
A decision matrix is moral with individual thresholds α_i iff. for all alternatives a, b:

$$\text{If } \exists \alpha_i \text{ s.t. } w_i v_i(a_i) > \alpha_i \text{ and } \sum_{i \in M} w_i v_i(a_i) > \sum_{i \in M} w_i v_i(b_i), \qquad (7)$$

then $v(a) > v(b)$

Reflects the way moral aspects of decisions are often evaluated in court, e.g. taking into account whether a crime has been committed out of a desperate need or sheer selfishness.
Problem: Makes moral attributes relevant that by the intended definition of a threshold are supposed to be irrelevant. This might confuse the decision maker and pose a problem for threshold elicitation.

Selective Thresholds View (Unipolar)

M_a^T: the set of indices i such that for alternative a, $w_i v_i(a_i) > \alpha_i$. A decision matrix is moral with selective thresholds α_i iff. for all alternatives a, b:

$$\text{If } M_a^T \neq \emptyset \text{ and } \sum_{i \in M_a^T} w_i v_i(a_i) > \sum_{i \in M_a^T} w_i v_i(b_i), \tag{8}$$

then $v(a) > v(b)$

Selective Thresholds View (Unipolar)

M_a^T: the set of indices i such that for alternative a, $w_i v_i(a_i) > \alpha_i$. A decision matrix is moral with selective thresholds α_i iff. for all alternatives a, b:

$$\text{If } M_a^T \neq \emptyset \text{ and } \sum_{i \in M_a^T} w_i v_i(a_i) > \sum_{i \in M_a^T} w_i v_i(b_i), \tag{8}$$

$$\text{then } v(a) > v(b)$$

This seems to be just the right condition. Only *relevant* moral attribute values are taken into account in a decision situation, but these need to dominate non-moral attributes.

Making the Account Bipolar (in a weak sense)

- So far, we have assumed that an attribute's value is morally relevant if it exceeds a given threshold.

Making the Account Bipolar (in a weak sense)

- So far, we have assumed that an attribute's value is morally relevant if it exceeds a given threshold.
- In practice, often the opposite is the case:

Making the Account Bipolar (in a weak sense)

- So far, we have assumed that an attribute's value is morally relevant if it exceeds a given threshold.
- In practice, often the opposite is the case:
    - lives lost

Making the Account Bipolar (in a weak sense)

- So far, we have assumed that an attribute's value is morally relevant if it exceeds a given threshold.
- In practice, often the opposite is the case:
    - lives lost
    - harm done

Making the Account Bipolar (in a weak sense)

- So far, we have assumed that an attribute's value is morally relevant if it exceeds a given threshold.
- In practice, often the opposite is the case:
  - lives lost
  - harm done
- Hence, we allow value functions to have negative values (as is usual) and introduce negative thresholds.

Bipolar Selective Thresholds View

M_a^T: the set of indices i such that

(Case 1) α_i is a positive threshold and $w_i v_i(a_i) > \alpha_i$, or

(Case 2) α_i is a negative threshold and $w_i v_i(a_i) < \alpha_i$

The previous Condition 8 remains unchanged; we sum over all attributes indexed by M_a^T:

$$\text{If } M_a^T \neq \emptyset \text{ and } \sum_{i \in M_a^T} w_i v_i(a_i) > \sum_{i \in M_a^T} w_i v_i(b_i),$$

then $v(a) > v(b)$

Simple Example With Only One Moral Attribute

Bob contemplates whether he should steal a chocolate bar from his cousin. For simplicity, all weights are 1 and the values are not normalized.

$M = \{1\}$, Threshold $\alpha_1 = -0.2$ (negative threshold)

	a	b
steal	-0.5	0.0
pleasure	0.8	0.2
Moral sum:	-0.5	**0.0**
Total sum:	**0.3**	0.2

Simple Example With Only One Moral Attribute

Bob contemplates whether he should steal a chocolate bar from his cousin. For simplicity, all weights are 1 and the values are not normalized.

$M = \{1\}$, Threshold $\alpha_1 = -0.2$ (negative threshold)

	a	b
steal	-0.5	0.0
pleasure	0.8	0.2
Moral sum:	-0.5	**0.0**
Total sum:	**0.3**	0.2

- $\Rightarrow$ Bob steals the chocolate bar. Bob's decision matrix is immoral.
- If stealing one chocolate bar would have a value equal to or above 0.2, then his choice would be moral (viz., morally excusable).

A Unipolar Example

$\alpha_1 = 0.1, \alpha_3 = 0.1$ (positive thresholds)

	a	b	c
1	0.7	0.2	0.4
2	0.3	0.7	0.3
3	0.2	0.5	0.7
Moral sum:	0.9	0.7	**1.1**
Total sum:	1.2	**1.4**	**1.4**

A Unipolar Example

$\alpha_1 = 0.1, \alpha_3 = 0.1$ (positive thresholds)

	a	b	c
1	0.7	0.2	0.4
2	0.3	0.7	0.3
3	0.2	0.5	0.7
Moral sum:	0.9	0.7	**1.1**
Total sum:	1.2	**1.4**	**1.4**

- Assuming condition 8 and the given thresholds, the decision matrix is immoral.
- However, actually choosing alternative *c* would be moral.
- This illustrates that the framework may express a virtue-ethical stance under certain circumstances.

Last Example

Loosely based on an example by von Wright (1963).

Situation: A cave diving expedition with three members goes horribly wrong. Two of Carl's colleagues, Alice and Bob, are gravely injured. Carl can only carry one of them back or he will likely perish himself during the rescue attempt. He is confident that the one he carries back will be saved. Whom should he save?

	a: Save Alice	b: Save Bob	c: Do nothing
1 – pers. pref.	0.50	0.25	0.25
2 – deaths	-0.50	-0.50	-1.00
Moral sum:	**-0.50**	**-0.50**	-1.00
Total sum:	**0.00**	-0.25	-0.75

Last Example

Loosely based on an example by von Wright (1963).

Situation: A cave diving expedition with three members goes horribly wrong. Two of Carl's colleagues, Alice and Bob, are gravely injured. Carl can only carry one of them back or he will likely perish himself during the rescue attempt. He is confident that the one he carries back will be saved. Whom should he save?

	a: Save Alice	b: Save Bob	c: Do nothing
1 – pers. pref.	0.50	0.25	0.25
2 – deaths	-0.50	-0.50	-1.00
Moral sum:	**-0.50**	**-0.50**	-1.00
Total sum:	**0.00**	-0.25	-0.75

$\Rightarrow$ The decision matrix is moral.

Conclusions

- Human decision makers often make errors. In light of the literature on actual human decision making and biases (Kahneman& Tversky), the framework is unlikely to be useful for retrospective assessments.

Conclusions

- Human decision makers often make errors. In light of the literature on actual human decision making and biases (Kahneman& Tversky), the framework is unlikely to be useful for retrospective assessments.
- Common value taxonomies seem to be compatible with the decision theoretic approach.

Conclusions

- Human decision makers often make errors. In light of the literature on actual human decision making and biases (Kahneman& Tversky), the framework is unlikely to be useful for retrospective assessments.
- Common value taxonomies seem to be compatible with the decision theoretic approach.
- The threshold view could be used to inform decision makers about moral aspects of their decisions without imposing overly pedantic requirements.

Conclusions

- Human decision makers often make errors. In light of the literature on actual human decision making and biases (Kahneman& Tversky), the framework is unlikely to be useful for retrospective assessments.
- Common value taxonomies seem to be compatible with the decision theoretic approach.
- The threshold view could be used to inform decision makers about moral aspects of their decisions without imposing overly pedantic requirements.
- In group decision making, moral attributes with thresholds could be identified by separate groups (‘ethics committees’) in addition to domain-experts. $\Leftarrow$ argumentation

Conclusions

- Human decision makers often make errors. In light of the literature on actual human decision making and biases (Kahneman& Tversky), the framework is unlikely to be useful for retrospective assessments.
- Common value taxonomies seem to be compatible with the decision theoretic approach.
- The threshold view could be used to inform decision makers about moral aspects of their decisions without imposing overly pedantic requirements.
- In group decision making, moral attributes with thresholds could be identified by separate groups ('ethics committees') in addition to domain-experts. $\Leftarrow$ argumentation
- Many open problems: non-additive models, incomplete preferences (cf. Hansson 2001), interval estimates, dealing with uncertainty in general, argumentative methods of preference elicitation, elicitation of moral thresholds and value functions, identifying moral attributes, social aspects, . . .

References

- Eisenführ, F., Weber, M. & Langer, T. (2010): *Rational Decision Making.* Springer.
- Fishburn, P. C. (1970): *Utility Theory for Decision Making.* John Wiley & Sons.
- Fishburn, P.C. (1991): Nontransitive additive conjoint measurement. *Journal of Mathematical Psychology*, Vol. 35, pp. 1-40.
- Gossen, H. H. (1854): *Entwicklung der Gesetze des menschlichen Verkehrs und der daraus flieenden Regeln fr menschliches Handeln.* Vieweg.
- Hansson, S. O. (2001): *The Structure of Values and Norms.* Cambridge UP.
- Jackendoff, R. (2006): The Peculiar Logic of Value.*Journal of Cognition and Culture*, 2006, 6, pp. 375-407.
- Jackendoff, R. (2007): *Language, Consciousness, Culture.* MIT Press.
- Kahneman, D. & Tversky, A. (1981): The framing of decisions and the psychology of choice. *Science*, vol. 211, no. 4481, pp. 453-458.
- Keeney, R. L. & Raiffa, H (1976): *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* John Wiley & Sons.
- von Neuman, J. & Morgenstern, O. (1947): *Theory of Games and Economic Behavior.* Princeton University Press.
- Von Wright, G. H. (1963): *The Varieties of Goodness.* Routledge & K. Paul.
- Savage (1954): *The Foundations of Statistics.* Wiley.