

Karl Dieter Opp (2013): Norms and Rationality: Is moral behavior a form of rational action?

Erich Rast

Nova Institute of Philosophy, Universidade Nova de Lisboa

Ethics and Political Philosophy Group, March 3, 2021



1 Introduction

2 Rational Choice

3 Defense of the Incentives Thesis

4 My Criticisms

Main Question

Can following *norms* be explained as being rational according to Rational Choice Theory (RCT)?

Definition

A norm is “[...] any statement claiming that something ought or ought not to be the case under certain conditions. A norm is *internalized* to the extent that following the norm is an intrinsic motivation or goal.” [384]

- “In this article, we focus on the internalization of norms [...]”
- So the question is really: Can the internalization of norms be explained by RCT? For instance, do rational agents follow norms because this helps them maximize their utility?
- morality \sim norm internalization

The article is concerned with morality only insofar as it is in place in a community to some extent.

Two Main Theses

Incentives Thesis

Social norms provide incentives just like material rewards do. Therefore, they can be integrated into Rational Choice Theory (RCT), and explanations of norm following can be based on RCT.

Autonomy Thesis

Following norms cannot be explained by RCT. People follow norms because they think they are right, because they have a duty or obligation to do so, irrespectively of their benefits. Obligations and duties cannot be explained as incentives.

Opp seeks to defend the Incentives Thesis against criticisms from proponents of the Autonomy Thesis.

According to Opp, RCT is characterized as follows:

- 1 Preferences: Preferences among future states are used for making a choice.
- 2 Beliefs: Constraints and behavioral opportunities determine behavior. In particular, an agent's beliefs are instrumental in making a choice based on preferences.
- 3 Utility Maximization: Rational agents maximize utility.

Desires are indirectly encoded by the preferences. This is more controversial in philosophy than Opp puts it.

Rational Choice Theory 2

“An action A is chosen that is believed (B) to be the best way to achieve P.”
[386]

RCT connects preferences (goals, desires) with beliefs (constraints) to result in actions.

External vs. Internal Outcomes

External Outcomes

“[...] the goal refers to phenomena outside the individual who performs the action.” [388]

Internal Outcomes

“This may be internal satisfaction if a goal is realized. [...] ‘Emotional disturbances’ (Rommelteit 1954, p. 58) are an internal outcome if the goal is not achieved. [...] If actions are *intrinsically rewarding* like jogging, this means that performing the action becomes a goal in itself.” [388]

Narrow Rational Choice Theory

Narrow RCT focuses on external outcomes only, i.e., material benefits and costs.

- 1 There is no restriction on preferences. They may be altruistic and include internal rewards.
- 2 Perceived constraints and beliefs matter. They may be biased or incomplete, for instance. (Simon 1991)
- 3 Utility is maximized from the perspective of the actor. It is not utility attributed by an omniscient 3rd party observer.

Example: Ultimatum Game

The Ultimatum Game

A dictator gets to split a certain amount $n + m$ of money between himself and a receiver. If the receiver rejects the offer, none of them get any money. If the receiver accepts, the dictator gets n and the receiver gets m .

- According to narrow RCT, it is rational for the receiver to always accept and rational for the dictator to maximize his amount, i.e., only give 1 cent to the receiver.
- According to wide RCT, it is rational for the receiver to reject an offer considered unfair. Fulfilling the presupposed fairness norm offers an internal reward that outweighs the monetary loss.
- In experiments, the amount is almost always split, sometimes up to 50:50.

Example: Paradox of Participation

Why do people vote?

According to narrow RCT, voting is irrational, since (i) it incurs a small cost and (ii) an individual vote does almost certainly not count. According to wide RCT, voting can be rational, since (i) it incurs a small cost but (ii) this cost is outweighed by the internal reward of fulfilling one's obligation as a citizen.

Elster (1989a): Outcome-orientation not compatible with norms

"Rationality says: If you want to achieve Y, do X. By contrast, I define social norms by the feature that they are not outcome-oriented. The simplest social norms are of the type: Do X, or: Don't do X...Rationality is essentially conditional and future-oriented. Social norms are either unconditional or, if conditional, are not future-oriented." [Elster 1989a: 99, cit. from Opp 2013: 390]

- Following a norm also has outcomes.
- Norms can also be conditional. (E.g. "Lie to someone if saying the truth would hurt that person")
- Elster's form "Do X" is compatible with the thesis that internalized norms are outcome-oriented. Following or breaking a norm can be described as being outcome-oriented if wide RCT is assumed.
- Internalized norms are also future-oriented. "The actor will then have the goal to fulfill the obligation the norm refers to whenever the respective situations will obtain." [392]
- Elster seems to assume narrow RCT without sufficient argument for it.

Etzioni (1986): The Problem of 'Mono-utility'

- RCT presumes one form of utility. Etzioni: pleasure and pain (referring to Bentham 1789)
- Acting out of duty and obligations is not based on pleasure and pain.
- Example: A woman dashes into a fire to save *her* child, the child she feels responsible for.

Remark (in my words): Etzioni's argument may be understood as pointing out that most utility theorists are value monists. There is one overall utility, and this value is hedonic. This seems to be at odds with the complexity of deontic moral concepts.

- 1a. Etzioni: satisfaction of pleasures \neq affirmation of moral commitments
- 1b. Opp: There are different kinds of satisfaction. Wide RCT can accommodate different kinds of utility.
- 2a. Etzioni: A broad conception of utility loses the ability to explain.
- 2b. Opp: Wide RCT is needed to make correct predictions (e.g. paradox of participation).

- 3a. Etzioni: Using the broad concept of utility is problematic because it is no longer possible to differentiate between the various kinds of motivations (like moral and non-moral motivations) that affect behavior. [paraphrase by Opp, 394]
- 3b. Opp: The theory does not differentiate, but the researcher can still differentiate between different kinds of motivations. (RCT idealizes from them.)
- 4a. Etzioni: A broad utility concept is unproductive and tautological. Whatever someone has done was done for pleasure. Cf. revealed preference thesis \rightsquigarrow circularity objection
- 4b. Opp: RCT can be tested. Multiple utilities can also be combined into overall judgments.

Remark: 4b is to some extent true but only if you elicit the preferences beforehand. Unfortunately, a lot of research in behavioral economics has shown violations of expected utility theory.

March & Olsen (2006): Logic of Appropriateness vs. Logic of Consequentiality

Logic of Appropriateness

This concerns appropriate or exemplary behavior, organized by institutions. Rules are followed because they are seen as rightful, expected, legitimate and embedded into social collectivity. [M& O: 690]

Logic of Consequentiality

The logic of consequentiality is concerned with self-interested and rationally calculating actors, instrumentalism and consequentialism. [M& O 2006: 691]

Can Wide RCT Accommodate the Logic of Appropriateness?

- The distinction between the two “logics” is not as clear-cut as March & Olsen seem to presume.
- In particular situations, the two logics may obtain.
- When *normative goals* conflict with *non-normative goals*, some compromise must be found.
- Opp argues that in such cases the actor will choose what's best for him. [398]
- An account of norm following not based on utility maximization must explain actors' motivation to follow norms.

Remark: Opp has already argued that people might not always choose what's best *for them*, so his remark on p. 398 is misleading. Note that best according to X \neq best for X.

Pro AT: Wrong Motivation Argument

"It is hardly plausible [...] that people follow norms in order to get a good conscience or in order to avoid a bad conscience, shame, or embarrassment. People act because they think that acting morally is right." [399]

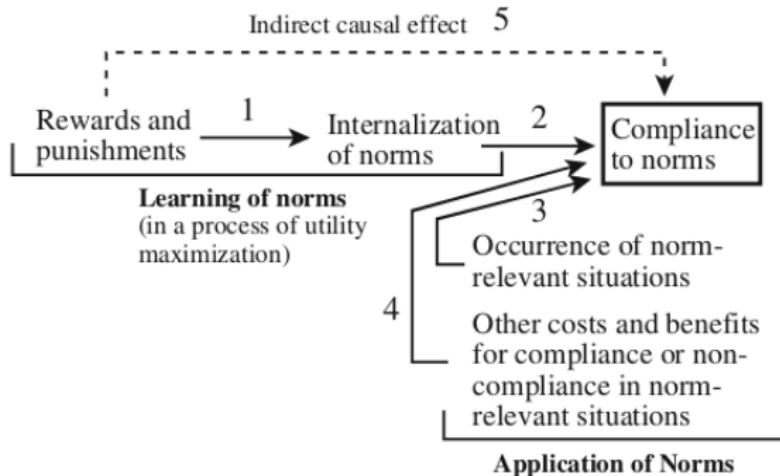
Opp's replies:

- The incentives thesis only stipulates that there is a goal to follow the norm if a norm is internalized. Feelings like good or bad conscience are only a by-product of such a process and need not be the decisive motive.
- Internalized norms provide their own incentives by following and costs by not following.
- In addition, other consequences of following or breaking the norm may play a role in explaining why actors follow or break internalized norms.

Norm Compliance Against Self-interest

- Examples: honor killings (Elster 1990); saving Jews from persecution by Nazis (Monroe et al. 1991)
- These are examples of risky behavior with extremely high costs.
- Opp's position: These examples do not show that utility maximization is not at play. The benefits could have been weighed appropriately high by the actors.

RCT Explanation of Norm Following



- Behavioral programs are learned on the basis of incentives, maximizing utility.
- Dual process theories: spontaneous mode versus deliberation.

Collective vs. Individual Benefits

- 1 Does complying with a norm increase utility?
- 2 Does the existence of a norm increase utility?

The second question regards the emergence of norms. Norms may e.g. evolve to reduce negative externalities, and there can be other group benefits. According to this view, norms increase individual utilities of group members. However, the original mechanisms may become defunct and the norm may stay in place because of the costs of changing the norm.

Criticism 1: Descriptive Inadequacy

- Opp's account is based on the idea of explaining norm following as utility maximization.
- There is plenty of evidence from behavioral economics and mathematical psychology that agents do not always maximize utility. See e.g. Lichtenstein & Slovic (2006).
- Example: cyclic preferences
- Example: prospect theory (anchoring point)

Criticism 2: Lack of Explanatory Power

- Etzioni's criticism might not have been answered sufficiently.
- Apart from structural constraints like the Expected Utility Hypothesis and acyclicity of underlying preferences, we can find matching preferences for any behavior.
- For the theory to have predictive power, preferences and utilities need to be elicited independently from the outcomes.
- Particular danger: Models in Evolutionary Game Theory with particular parameter choices. There needs to be independent evidence for these parameters.

Criticism 3: Unclear Relation between Preferences and Incentives

- If X prefers A over B, and there is a choice between A and B, then X will choose A.
- Benefits, rewards, incentives are not needed in this account.
- They seem to be part of some 'psychological explanation' for following preferences (motivation).
- But if incentives etc. existed (early utilitarianism!), they could replace preferences.
- If preferences are the basis in the first place, then the whole debate about internal vs. external rewards and wide vs. narrow RCT becomes spurious.

To conclude, RCT can explain any behavior consistent with preference relations. However, it is doubtful whether the RCT description has any benefits for the explanation of norm following over other explanations, unless individuals mention preferences (or, benefits, incentives, etc.) as the *reasons* for acting that way in surveys.

References

- Elster, J. (1989a). Social norms and economic theory. *Journal of Economic Perspectives*, 3, 99–117.
- Elster, J. (1989b). *The cement of society. A study of social order*. Cambridge: Cambridge University Press.
- Etzioni, A. (1986). The case for a multiple-utility conception. *Economics and Philosophy*, 2, 159–183.
- Etzioni, A. (1988). *The moral dimension. Toward a new economics*. New York: Free Press.
- Lichtenstein, S. & Slovic, P., eds. (2006). *The construction of preference*. Cambridge UP.
- March, J. G., & Olsen, J. P. (2006). The logic of appropriateness. In M. Moran, M. Rein, & R. E. Goodin (Eds.), *The Oxford handbook of public policy* (pp. 689–709). Oxford: Oxford University Press.
- Opp, K.-D. (2013). Norms and rationality. Is moral behavior a form of rational action? *Theory and Decision*, Vol. 74, No. 3, 383-409.