# Instrumental Rationality in the Social Sciences

*Katharina Nieswandt, Concordia University, Montreal*

August 5, 2023

## Publication & Download

*Philosophy of the Social Sciences*, 2023, online first.
DOI: 10.1177/00483931231181930.

## Key Words

consequentialism * rational choice theory * backward-looking motive * revenge * disgust * logical ghost * anthropology * psychology

## Acknowledgments

I am deeply grateful to the many colleagues who commented on earlier versions of this paper at various conferences. I am particularly indebted to Ulf Hlobil, Tristan Rogers, and two anonymous reviewers.

## Abstract

This paper draws some bold conclusions from modest premises. My topic is an old one, the Neohumean view of practical rationality. First, I show that this view consists of two independent claims, instrumentalism and subjectivism. Most critics run these together. Instrumentalism is entailed by many theories beyond Neohumeanism, viz. by any theory that says rational actions maximize something. Second, I give a new argument against instrumentalism, using simple counterexamples. This argument systematically undermines consequentialism and rational choice theory, I show, using detailed examples of their many social science applications. There is no obvious fix.

# 1 A surprising gap in the literature

Practical rationality is often seen as the ability to reason from some adequate end to adequate means. On this view, all rational actions are instrumental actions. Let's call it "instrumentalism."

**Instrumentalism:**

> An action is rational iff it is an adequate means to an adequate end.

Instrumentalism is dominant in many corners of philosophy, in the social sciences and even in common sense. Philosophers credit David Hume (2000, sect. 2.3.4 and 3.1.1) with developing, or at least popularizing, instrumentalism. Hence many speak of the "Humean theory of practical rationality." Common labels for contemporary variants are "Neohumeanism," "belief-desire theories" and "noncognitivism."

Hume scholars tend to be critical of this attribution.[1] People outside of philosophy have, anyway, rarely heard of Hume but still hold the view. Indeed, instrumentalism is so dominant that many treat it as a conceptual truth rather than a theory.

Within metaethics, instrumentalism has been criticized strongly, at least since the 1950s.[2] The debate, however, is largely confined to metaethics. I am not aware of any systematic analysis of what other theories include instrumentalism and are hence affected by the criticisms.[3]

This gap is striking because "instrumentalist theories," as we may call them, are numerous. Any theory on which a rational action is one that maximizes some outcome, I shall argue (section 2), includes instrumentalism. Important examples are rational choice theory and

[1] For comprehensive critiques, see Cohon (2008) and Radcliffe (2018).

[2] Critics argue that the end itself must also be rational, at least in the sense of intelligible, if the means towards it are to be rational (Quinn, 1993), that beliefs alone can motivate us (Korsgaard, 1986), that ends can be objectively reasonable or unreasonable (Foot, 2001, ch. 4), that the end is the object being desired rather than the desire fulfillment (Anscombe, 2000, ch. 34) and that desires are not causally efficacious in the relevant sense (Nagel, 1978, ch. 5)—to give just a few examples.

[3] One of the few existing explorations is Teichmann (2018).

consequentialism, as well as their many applications in the social sciences. Among the latter are the *homo economicus* paradigm in economics, theories of vendetta (Chagnon, 1988; 2012, ch. 6) and of gifting in anthropology (Malinowski 2002, chs. 1.3–4 and 1.8–9; Firth, 2011, ch. 12; Sahlins, 2017, chs. 4–5), and certain theories of social intelligence (Weinstein, 1969; Goleman, 1998) and of moral justification in psychology (Haidt, Björklund, and Murphy, 2000; Haidt, 2001; Haidt and Hersh, 2001).

My paper fills this gap. My key idea is that there are obvious examples of noninstrumental but rational actions (sections 3–4). These cannot be rejected no matter what theory of practical rationality we presuppose, but would have to somehow be accommodated within any such theory. Even more, such actions are ubiquitous. Rational choice theory and consequentialism (section 5), I then show, as well as their social science applications (section 6), cannot properly accommodate such actions; they systematically misinterpret them.

These social science applications are, in fact, helpful illustrations of what is wrong with the instrumentalist conception of rational actions. It is too narrow. Neither rational choice theory nor consequentialism, however, can easily be amended so as to broaden their conception (section 7). They are, therefore, unable to make sense of a large spectrum of human actions, which ranges from avenging to helping one's friends.

## 2 Instrumentalism versus subjectivism

To understand my argument, it is helpful to understand the reasons for the gap. The gap exists, I believe, because many critics of Neohumianism conflate problems for instrumentalism with problems for subjectivism and because their arguments effectively criticize subjectivism. "Neohumeans," as I shall use the label here, conceive of rational actions as actions suitable to achieve some end *(instrumentalism)*. Additionally, they conceive of ends as fully and exclusively justified

by the agent's desires *(subjectivism)*. They hence hold to two claims.

**Neohumeanism:**

> An action is rational iff it is an adequate means to an adequate end. *(instrumentalism)* &
>
> An end is adequate iff the agent desires it. *(subjectivism)*

These two claims are logically independent of one another. Suppose, I believe that donating my salary to an orphanage is an adequate means to the adequate end of alleviating child poverty. What would I, additionally, need to believe about the end if my beliefs are to be consistent? A Neohumean would additionally believe that what makes the alleviation of child poverty an adequate end is that I desire to do so. A classical utilitarian, however, would instead believe that what makes the alleviation of child poverty an adequate end is that it enhances the global pleasure-pain balance of all sentient beings. Whether I desire to do so or not does not matter for the question whether the end is adequate. A classical utilitarian hence is an instrumentalist without being a subjectivist—and without being inconsistent.

Neohumeans are both instrumentalists and subjectivists, so their critics often fail to separate the two claims. This includes the critique by Thomas Nagel (1978, ch. 5) I later discuss (section 5). Candace Vogler's (2002, p. 2) book-length investigation into the true core of instrumentalism even defines "instrumentalism" so that subjectivism is a constitutive part of it. Other critics, while not conflating the two, focus on subjectivism. This includes Christine Korsgaard (1986), Warren Quinn (1993) and Philippa Foot (2001).

It takes a small, but necessary philosophical step to see that some of the arguments levelled against Neohumeanism concern the first above claim only. Therefore, they apply to any theory which includes this first claim, rather than only the Neohumean package deal. Whether a theory understands ends as "passions," as "preferences" in some minimal sense or as impersonal "good states of the world" does not matter for these arguments.

Once we take this step, an interesting implication is obvious. Problems for instrumentalism are problems for any theory that includes instrumentalism as a central component, whether subjectivist or not. They cannot be evaded by allowing the evaluation of ends as rational or irrational.

On this minimalist understanding of instrumentalism, there are many instrumentalist theories, viz. any theory that justifies actions exclusively by their outcomes. Rational choice theory and consequentialism are the most influential theories of this kind. Today, there are so many different variants of each that it seems accurate to regard them as families rather than individual theories. The minimally shared commitment of all, however, is that a rational action is one that maximizes (expected) outcomes—where outcomes are defined as the satisfaction of agential preferences and as non-agent-relative good consequences, respectively.[4] Both hence regard a rational action as a means to an end, either maximal preference satisfaction or maximally good consequences. Both will hence be affected by problems for instrumentalism.

The impact of our small discovery, however, extends even beyond this. While rational choice theory and consequentialism loom so large today as to constitute families of theories, many standard theories in disciplines beyond philosophy entail instrumentalism as a central commitment. These instrumentalist theories, too, should be affected. I hence supply two short case studies that illustrate how.

## 3 My argument

I shall not attempt a comprehensive review of extant arguments against instrumentalism and how each of them applies to rational choice theory and consequentialism, respectively, as well as to their many uses in empirical research. That would be impossible even in a book-length treatment, and it would also be very repetitive.

[4]See, e.g., Ross (2019, sect. 2.1) and Sinnott-Armstrong (2019, sect. 2). You might be wondering about indirect consequentialism here; I address this in section 7.

Instead, I develop one argument in detail (sections 4–5) and show how it applies to one social science use each of rational choice theory and of consequentialism (section 6). A detailed discussion of this one point will illustrate the systematic problem with instrumentalism well enough. As a last step, I consider some obvious defenses for rational choice theorists and consequentialists, arguing that none succeeds (section 7).

I hope to pull a fat rabbit out of a tiny hat here. My argument does not require you to subscribe to a specific framework in action theory, ethics or philosophy of science. I remain agnostic regarding the correct theory of virtually all major topics I discuss, from rationality to revenge to altruism. All I attempt to show is that a certain view of practical rationality cannot be correct because it faces obvious counterexamples and cannot easily be amended so as to incorporate these.

My argument, in other words, makes minimal philosophical presuppositions. It looks as follows.

1. Some rational actions are not instrumental. *(section 4)*
2. Instrumentalist theories misinterpret such actions.
3. Rational choice theory and consequentalism are instrumentalist theories, therefore misinterpret such actions. *(section 5)*
4. This shows in their social science applications. *(section 6)*

Of these four claims, the first is the point of contention. It denies what instrumentalism says, viz. that all actions that are rational also are instrumental. I take some inspiration from Elizabeth Anscombe (1981; 2000, chs. 12–14) to argue for it, in section 4, but you do not have to accept her theory of action in order to accept my argument because I proceed by means of simple counterexample. My cases of actions from motives such as revenge, gratitude, kinship or solidarity will be acceptable to readers from diverse philosophical backgrounds. Some readers will reject individual examples, I expect, but it would be a serious philosophical bullet to bite to reject them all.

Indeed, instrumentalist theories, I proceed to show, usually choose a different strategy. They accept these actions as rational and attempt to reinterpret them in an instrumentalist fashion. These reinterpretations, we shall see, cannot succeed; they clearly distort what belongs to a genuinely different category of actions. Rational choice and consequentialist reinterpretations of actions such as avenging, in other words, are misinterpretations rather than simply legitimate recastings in a different vocabulary.

Last, I move on to two social science examples in section 6. One is a classic from anthropology: Napoleon Chagnon's (1988) groundbreaking analysis of vendetta, which uses rational choice theory. My other example is a recent use of consequentialism in psychology and experimental philosophy: Jonathan Haidt's (2000; 2012) analysis of our supposed inability to rationally justify disgust, as evidenced by an alleged "moral dumbfounding" of study participants. These examples are more than mere illustrations of a general claim already established at that point. They provide insight into the root of the problem, uncovering what exactly is wrong with instrumentalism. The instrumentalist view of practical rationality is, if not simply false, at least much too narrow.

## 4 Non-instrumental rational actions

My first claim (from page 7) says that there are non-instrumental rational actions. An instrumental action is one that is carried out as a means to an end (see my definition of instrumentalism on page 3). You act *in order to* achieve $A$. That which justifies your action, if it does, i.e. which provides a good reason for your action, is its expected consequence. By definition, a consequence lies in the future. E.g., you might give Jack money today in order that he take over your shift next week.

Many actions, if you think about it, do not follow this pattern. Take actions from gratitude or revenge. In both cases, that which justifies the action (if it does) is a past event. You act, not in order to achieve

| **Motive** | **Reason is in the** | **You act...** | **Examples** |
|---|---|---|---|
| *instrumental/ forward-looking* | future | ... in order to | material gain, health, security |
| *backward-looking* | past | ... given that earlier | gratitude, revenge, fairness |
| *interpretative* | background | ... since | friend-/kinship, solidarity, disgust |

Table 1: Three kinds of rational actions

*A*, but *given that earlier B* happened. You buy Jack a drink today given that, last week, he helped you acquire a permit; or you kill Jack today given that, last week, he killed your brother.

Anscombe (1981; 2000, chs. 12–14) calls the first kind of action, my "instrumental" actions, actions from "forward-looking" motives and the second kind actions from "backward-looking" motives. She also recognizes a third kind, actions from "interpretative" (or "general") motives. She describes the last kind of motive as one that asks you to "See the action in this light" (2000, p. 21). Thus, you might buy Jack a drink *since* he is your friend, or you might refrain from killing Jack *since* he is your son.

Table 1 depicts the three kinds of (potentially) rational actions. Let me preface my discussion of it with four clarifications.

First, it is possible to doubt for some of the listed actions that they can ever be rational. Revenge, e.g., may never be rational. Second, some of my examples can be placed into another category, depending on the exact situation. "Fairness," e.g., is a backward-looking motive if you act given that the others already did earlier but would be interpretative if you act since all should receive the same. I.e., you can mean "fair" in the sense of 'reciprocal' or in the sense of 'equal'. Third, actions can have mixed motives or be overdetermined. I might buy Jack a drink both from gratitude and because he is my friend. Fourth, ac-

tions in the last two rows may consist in many steps, several of which can be described in a forward-looking manner. The action of avenging, e.g., could involve the purchase of ammunition in order to shoot Jack.

The point of table 1 is to illustrate that other kinds of rational actions than instrumental actions exist indeed. *Any* example that could correctly be placed in the last two rows supports my first claim.

In order to challenge this claim, you'd either need to say that all actions in the last two rows are irrational or that all actions in these two rows are instrumental actions in disguise.

The first reply is so implausible as not to merit discussion. Some will challenge some of the above motives, and many will challenge particular instantiations. As indicated, some will hold that it always is irrational to take revenge, and many will hold that the particular action of murdering Jack is irrational, that the rational choice in our kind of society is to take the matter to the authorities. Few, however, will want to claim that all these motives are irrational, hence any imaginable instantiation is. Few, in other words, will say that it can never be rational to be grateful, be fair or show solidarity. Surely, it is not always irrational to write a thank-you card, reciprocate favors or sign the petition to free a political prisoner? These are, furthermore, only three random examples; we could easily generate others. I therefore proceed to a detailed look at the second reply.

The second reply uses a strategy called "consequentializing"—we could also call it "instrumentalizing" here. There are two ways to instrumentalize an action.

The first version of this reply reinterprets the action's motive as forward-looking. Perhaps the reason why you buy Jack a drink is that you want him to help you again, in the future.[5] Call this the "Some more please!" theory of gratitude.[6] Similarly, the reason why you kill

[5]There is a collectivist version of this (Mauss, 2016), which appears more plausible than the version for individual agents, the anthropologist paradigm of the "gift economy."

[6]Anselm Müller once suggested this label to me.

Jack might be that you want to deter others from attacking your kin. The latter view is a "deterrence theory of punishment," a family of theories with numerous proponents.[7]

The problem with such reinterpretations of the motive as instrumental is that they effectively declare the action to be a different kind of action. If you buy Jack a drink solely, or even mainly, so that he help you again, then what you do is not to thank Jack—you bribe him, or you prepare a future business opportunity, or ...., whatever the concrete circumstances. I.e., the very same movements, utterances etc. (such as those involved in the buying of the drink) from a different motive would constitute a different kind of action. One way in which this comes out is that the appropriate reaction often is disappointment if someone discovers that an action seemingly done from a backward-looking or interpretative motive truly was forward-looking: "Oh, you did that just so you then ...." Jack, considering himself your friend, would rightly be disappointed to hear that, all these years, you only went drinking with him because he works at the planning office.

This point has been discussed extensively for deterrence theories of punishment. In 1764, Cesare Beccaria (1995, ch. 12) synthesized various enlightenment ideas on utility and state authority into the theory that the sole end of punishment is prevention through deterrence, with an emphasis on general rather than special prevention. Jeremy Bentham (1996) took the next logical step by generalizing Beccaria's legal into a moral perspective, developing two tenets that characterize "consequentialism," i.e. the broad school of normative ethics, until this day: the reinterpretation of concepts such as 'innocence', 'responsibility' or 'desert' in terms of utility and the claim that consequentialism is rational precisely because of this reinterpretation. Thus, the steady stream of *reductio* arguments that consequentialism has produced since its introduction—such as the argument that consequentialism might require us to 'punish' the innocent[8]—are usually answered by claiming that the fundamental revision of common moral

[7] Recent defenses are by Tadros (2011) and Ellis (2003).

[8] E.g., McCloskey (1957, pp. 468–69) or Anscombe (1981b, pp. 39–42).

principles which the theory requires in fact exposes how irrational these principles were in the first place.[9]

The revisionist reply, however, cannot possibly satisfy the opponent—as the much subtler John Stuart Mill (1985, ch. 5) realizes. The accusation is not that consequentialism has difficulties incorporating this-or-that principle or counterexample; it is that consequentialism gives up the very concept of justice rather than correct our understanding of it. Another way to put the accusation would be to say that justice as relating to punishment and reward is a backward-looking motive, but that consequentialists reinterpret it as the forward-looking motive of utility and thereby instrumentalize the action done from it, thus declaring it to be a different kind of action rather than offering a better justification for the same action.

Notice that it would not help to identify the good outcome with the action itself. We could imagine a consequentialism that instrumentalizes actions by assigning a value to the action itself and by then declaring the production of this value the reason for the action. On this view, I should buy Jack a drink, e.g., because I should do whatever creates the best possible world, and a world with this act of gratitude in it is better than the alternative world without it. Similarly, I should kill my brother's killer because revenge is valuable in and of itself; hence the world is better for this act of revenge than without it.

This imaginary consequentialism, too, would turn the motive into a forward-looking one and would hence declare the action to be of a different kind. If I kill my brother's killer to produce some outcome unrelated to my brother and his death, or if I buy Jack a drink to produce some outcome unrelated to the help Jack gave me, then what I do is neither avenging nor thanking. One indication of this is that, on this picture, I could carry out any random other action, for instance donate a thousand dollars to an orphanage, provided it creates the same amount of goodness, and my brother's murder or Jack's help would be an equally valid reason for this random other action.

[9] E.g., Singer (1972, pp. 233–35) or Harris (1975, p. 83).

The first way to instrumentalize an action hence lacks appeal for all who are not already committed to consequentialism for independent reasons. The second way is more subtle and more promising. You can also instrumentalize an action by declaring its goal-directedness a form of instrumental justification. Since all actions are goal-directed at least in a minimal sense, you have thus instrumentalized all actions!

Anscombe (2000, chs. 20–22) considers and rejects this idea. Her argument went unnoticed in 1957, but a closely related argument by Nagel in 1970 has been extensively discussed under the heading of "logical ghosts." The two arguments are complementary, I believe, and I will put them together in section 5. To refute the second version of the second reply, however, Anscombe's argument alone suffices. The argument is this:

> [I]f I kill a man as an act of revenge I may say I do it *in order to* be revenged, or that revenge is my object; but revenge is not some further thing obtained by killing him, it is rather that killing him *is* revenge. Asked why I kill him, I reply "Because he kill*ed* my brother." (2000, p. 20, emphases added)

The mere logical fact that, qua action, any action has a goal is not, Anscombe points out here, enough to classify all actions as instrumental. An action is instrumental only if it is a means to an end; again, see my definition of instrumentalism on page 3. That end cannot be identical with the means towards it, i.e., with the action itself. If I act *in order to,* then my action and that in order for which I act are two distinct entities.

Readers might wonder whether there couldn't be cases in which means are constitutive of their end. If I find that scuba diving is fun, e.g., then isn't the action of diving my means to the end of having fun while that end, too, consists in the diving and hence the action itself? And can't we nevertheless logically separate the two, given that the same action might not produce the outcome of fun for other people and that other actions than diving could produce the same outcome of

fun for me?

This objection, however, trades on an equivocation of "fun" as an outcome that can be separated from the action versus "fun" as integral part of an action. On the first understanding, you separate the movements I conduct in diving from, e.g., the sensory stimulation I experience while moving (or from whatever you believe "fun" consist in), and you call the latter "fun." If that is your picture, then diving (and other fun activities) cannot serve as your example because you regard it as a standard case of instrumental action. I conduct the movements in order to trigger the stimulation; the movement is a means, and the stimulation is an outcome; there is no constitutive relation.

On the second understanding, "fun" is an aspect of the action itself. The movements, the sensory stimulation etc. together make up the action of diving. In that case, however, fun activities cannot serve as your example either—now, because you do not regard them as means at all; there *only* is a constitutive relation. On this picture, saying that "I dive in order to have fun" is like saying that "I dive in order to be diving" or that "I have fun (in this particular way) in order to be having fun (in general)." Similarly, in Anscombe's original example, "revenge is not some further thing obtained by killing him, [...] killing him is revenge." Actions are always "under a description," as Anscombe would call it, and "I dive in order to have fun" switches the description halfway through, from "diving" to "my preferred way of having fun," and while that is perfectly legitimate as a description, it does not license the metaphysical inference that the first part of the sentence describes a means for an end then described in the second part of the sentence.

Hence, unless there is reason to reject my initial definition of instrumental actions, we cannot instrumentalize actions by pointing to their goal-directedness. Table 2 on the next page summarizes all discussed replies to my first claim and their problems again.

What I said in this section is true not only of actions from backward-looking but also from interpretative motives. If I choose not to kill the killer after all because he is my son, then my reason is kinship. His

| | Reply | | Version | Problem |
|---|---|---|---|---|
| 1 | All non-instrumental actions are irrational. | | — | Many counter-examples (e.g., thanking). |
| 2 | All non-instrumental actions, in truth, are instrumental. | 2.1 | The motive is instrumental. | The action changes (e.g., from thanking to bribing). |
| | | 2.2 | Goal-directedness is instrumentality. | Contradicts definition of instrumental action. |

Table 2: Instrumentalizing and its problems

being-my-son is the reason why I forgo revenge; the reason is not some further thing obtained by not killing him, such as a future in which my son is still alive. As with the backward-looking reason of revenge, my action of sparing the killer's life (or: my omission to kill him) *is* his survival.

Hence, it seems that there are actions, including at least some of my examples from table 1, that are rational but are not carried out as a means to an end. These actions, I have attempted to demonstrate, cannot plausibly be reinterpreted as means to an end, neither by tracing them to a different motive nor by equating goal-directedness with a means-end relation. They are of a genuinely different kind. I conclude that some rational actions are not instrumental and that instrumentalist theories misinterpret such actions.

# 5 The full picture

I hope to have made a strong case for my first and second claim. Rational choice theory and consequentialism, we already gathered from the examples in section 4, are instrumentalist theories. Put in Anscombe's terminology, both accept only forward-looking motives. Let me show

how exactly problems arise for each theory in dealing with actions from backward-looking and interpretative motives.

Earlier, I mentioned that Nagel (1978, ch. 5) makes an argument which complements Anscombe's, and it is through putting their two arguments together that the full picture emerges. Nagel says:

> [I]t is true [...] that *whatever* may be the motivation for someone's intentional pursuit of a goal, it becomes in virtue of his pursuit *ipso facto* appropriate to ascribe to him a desire for that goal. But if the desire is a motivated one, the explanation of it will be the same as the explanation of his pursuit [...]. [...] [N]othing follows about the role of the desire as a condition contributing to the motivational efficacy of those [i.e., the agent's] considerations. It is a necessary condition of their efficacy to be sure, but only a logically necessary condition. It is not necessary either as a contributing influence or as a causal condition. (1978, pp. 29–30, emphases original)

Thus, it would be correct to say that, in killing the killer, I display a desire to kill him, which killing him fulfills. Nothing, however, follows about the role of my desire as a "contributing influence." In fact, my action of killing and my desire to kill the original killer are explained by one and the same thing, viz. him killing my brother. They are both "motivated" by the same entity (1978, p. 28) and might directly originate from the same source, so the desire cannot be shown to cause or in any other way contribute to the action. According to Nagel, this source is a reason (1978, p. 30). Anscombe would say that the answer to the question "Why?" is the same for both the desire and the action. I desired to kill him and I killed him—why?— both because he killed my brother.

Like many critics (see section 2), Nagel addresses his argument not only to instrumentalism but also to the subjectivism that Neohumeans combine with the former. We can substitute his term "desire" for "preference," however, and suddenly the argument applies to ra-

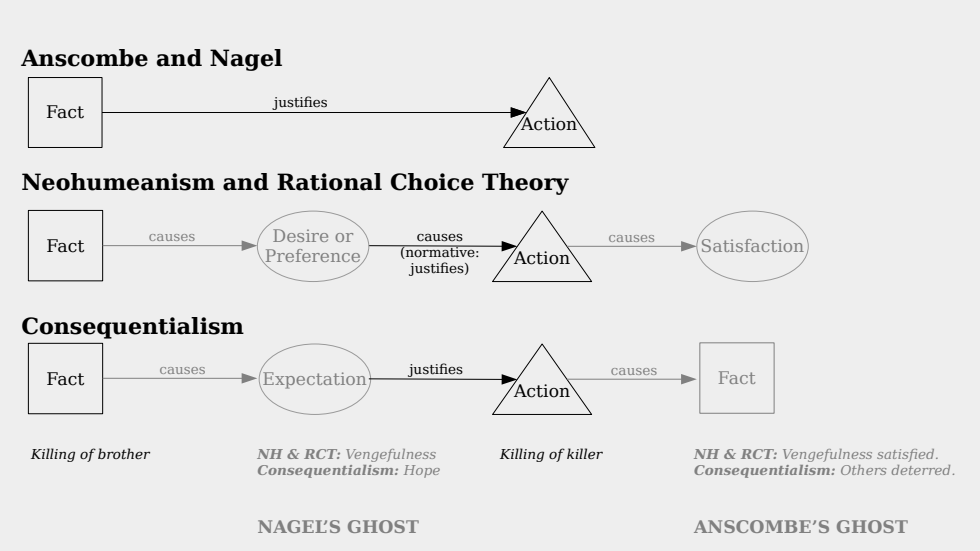


Figure 1: Two logical ghosts.

tional choice theory, whereas if we substitute it for "expectation," it applies to consequentialism. His relevant philosophical point, in other words, concerns instrumentalism alone. Figure 1 illustrates how the two arguments affect each theory and how they complement one another.

Anscombe and Nagel have a simple, two-step picture. Some fact—by which I mean everything from a personal relation to a datable event—justifies an action, provided the action is indeed rational. Neohumeans insert two additional steps. The fact causes a desire in the agent, which then causes an action, which satisfies the desire. Rational choice theorists share this picture, except that they substitute 'desire' for the wider concept of a preference. Both theories can be purely descriptive, but they also come in a normative variant, on which desires or preferences do not (only) cause but (also) justify the action.[10]

This four-step picture, so my combined argument, inserts two logical ghosts. According to Nagel, the desire (or preference) for revenge may be a logical ghost of the action of avenging and so is, Anscombe (2000, p. 20) adds, the desire's fulfillment because "revenge is not some

[10]See Gauthier (1992) and Briggs (2019).

further thing obtained by killing." The desire to kill the killer, i.e. the desire to take revenge, if we want to postulate such a thing, is fulfilled by killing him; in other words, the desire fulfillment and the action fall into one.

Subjectivist consequentialists (i.e., the vast majority of contemporary consequentialists) also have a four-step picture, but they insert other ghosts. They do not postulate a desire; instead, the agent has the expectation that a certain action will produce some beneficial fact. This expectation justifies the action—again, provided the action is indeed rational. For the example of revenge, as discussed in section 4, many consequentialists hold that this is the expectation of deterrence. The action then causes the expected beneficial fact.

Objectivist consequentialists skip step 2. They thus avoid Nagel's ghost, but not Anscombe's.

I conclude that rational choice theory and consequentialism are instrumentalist theories indeed, therefore misinterpret non-instrumental rational actions.

## 6 Social science examples

The two examples to be discussed in this section were selected because they are influential, within their fields and beyond, and because they are maximally diverse with respect to our topic. They come from two different fields; the first uses rational choice theory to explain an action from a backward-looking motive, and the second uses consequentialism to explain an action from an interpretative motive. They hence furnish ideal illustrations of how the instrumentalist paradigm shapes social science.

My first example is Napoleon Chagnon's (1988; 2012, ch. 6) analysis of vendetta among the Amazonian Yanomami, an application of rational choice theory in anthropology. This was so influential that the textbook went through six editions, the most recent subtitled "legacy edition," and it also became a public bestseller. Chagnon lived among

the very isolated Yanonami for extended periods of time and initially struggled to make sense of what he depicted as a never-ending cycle of revenge killings that imposed severe costs on all involved parties.[11] Rational choice theory seemed to offer the solution.

> The most common explanation given for raids (warfare) is revenge *(no yuwo)* for a previous killing [...].
>
> *At first glance*, raids motivated by revenge *seem counterproductive*. Raiders may inflict deaths on their enemies, but by so doing make themselves and kin prime targets for retaliation. But ethnographic evidence suggests that revenge has an *underlying rationality*: swift retaliation in kind serves as a deterrent over the long run. War motivated by revenge seems to be a tit-for-tat strategy in which the participants' score might best be measured in terms of minimizing losses rather than in terms of maximizing gains.
>
> [...] Losing a close genetic relative (for example, a parent, sibling, or child) potentially constitutes a significant loss to one's inclusive fitness. Anything that counterbalances these losses would be advantageous. (1988, p. 986, emphases added)

Notice that Chagnon's motivation in this passage is to justify the Yanomami's actions—not in the sense that he advocates vendetta, of course, but in the sense that he seeks a description of vendetta as rational in the sense of producing an overall beneficial outcome for the agent.[12] I.e., we here have an application of normative rational choice theory. The described agents themselves justify their action by "a previous killing," i.e., by a past event. Their justification is backward-looking. Chagnon regards the action *thus described* as irrational. His remedy is to reinterpret it as forward-looking. In other words, Chagnon instrumentalizes the action, purporting to thereby uncover its "underlying

[11] Chagnon (2012) referred to the Yanomami as "the fierce people." Others have doubted his rendering of the facts (e.g., Albert, 1989; Ferguson, 2001).

[12] For a philosophical defense of this kind of view, see Elster (1990).

rationality." Once the action is instrumental/forward-looking, it supposedly is rational.

Chagnon's 'rationalization' of Yanomami vendetta epitomizes the idea that a rational action must be a means to an end. Against this, I argued in section 4 that there are non-instrumental rational actions and that revenge, if it is rational, is justified precisely by the reason that the agents volunteer here. "Asked why I kill him, I reply 'Because he killed my brother'" (Anscombe, 2000, p. 20). Note that my argument applies to all items Chagnon wants to include in his instrumentalization, viz. the avenger's gains, costs and opportunity costs, both immediate and long-term, as well as both direct, as in inflicting costs on the target, and indirect, as in tilting the overall genetic balance in one's favor. The argument, in other words, applies to every future end to be brought about by the killing.[13]

Attempts to uncover the underlying rationality of revenge by instrumentalizing it hence are unnecessary, but they also misinterpret the observed behavior. The very impression that "revenge seem[s] counterproductive" proceeds from the premise that any rational action must produce something, which then leads to the question what revenge produces, which Chagnon answers with "inclusive fitness." As I argued in section 4, this premise is false. Once we drop it, Chagnon's puzzle vanishes. The Yanomami's killing of the killer or his kin can be understood for what it probably is, viz. the punishment of a person who wronged you or your kin. Thus understood, the fact that the action is costly does not pose a puzzle—or at least none that is specific to the Yanomami.

My second example is from contemporary psychology. Jonathan Haidt, social psychologist and influential public speaker, faces a similar puzzle as Chagnon. Starting with his doctoral dissertation, he conducted multiple studies on disgust and found that participants display a "stubborn and puzzled maintenance of a judgment [that an

[13]Interestingly, Chagnon's claim even has insufficient empirical support. His results (1988, pp. 989–90) by no means establish that vengefulness increases genetic fitness—or even correlates with it. (Chagnon provides no inferential statistics.)

action is disgusting] without supporting reasons," a reaction which he "dubbed 'moral dumbfounding'" (Haidt, Björklund, and Murphy, 2000, abstract). Haidt's puzzle, we shall see, only arises because his experiments apply consequentialism to an action (or rather, a reaction) from an interpretative motive.

The following is Haidt's best-known experiment.

> [W]e used two '[moral] intuition' stories, written to be simultaneously *harmless yet disgusting*. One of these stories (Incest) depicts consensual incest between two adult siblings, and the other (Cannibal) depicts a woman cooking and eating a piece of flesh from a human cadaver donated for research to the medical school pathology lab at which she works. These stories were chosen because they were expected to cause the participants to come to a quick and intuitive 'seeing-that' the act described was *morally wrong. Yet* since the stories were carefully written so that nobody in them was harmed, *participants are prevented from engaging in the usual 'reasoning-why'* that persons in Western cultures often use to justify moral condemnation [...]. (Haidt et al. 2000, p. 6, emphases added)

Participants read these vignettes of "harmless yet disgusting" actions and then answered the question: "Was there anything wrong with what [s/]he did?" Those who affirmed had to defend their judgment against the experimenter, who insisted that what does not result in actual harm cannot be wrong. I.e., we here have an application of objectivist act-consequentialism.

Haidt claims to have observed, in this and similar studies (Haidt and Hersh, 2001; Haidt, 2012, pp. 3–4, 95), that "a strong intuition is left unsupported by any reasons that can be verbalized" (Haidt et al. 2000, p. 13). I.e., participants are "dumbfounded," but that does not move them to change their judgment.

I would go so far as to call Haidt's results an artifact of measure-

ment.[14] He (implicitly) presupposes that the only reason to judge an action wrong is that it actually inflicts harm. He then presents cases where, by stipulation, no harm is being inflicted and finds that participants cannot point to the harm being inflicted. In other words, the only reason why a puzzle arises in Haidt's experiments is his consequentialist framework.

Haidt shares Chagnon's premise that only instrumental actions are rational actions. Contrary to Chagnon, however, he does not attempt to rationalize the observed behavior by instrumentalizing it. Instead, he declares it to be irrational. In fact, Haidt reads his results as an empirical confirmation of Hume's theory of practical rationality. He interprets disgust as a feeling and hence beyond the purview of practical rationality.

> The present study, however, tests *Hume's claims* empirically. [...] [...][We] investigate a class of moral dilemmas in which reason and passion conflict. If Hume is (generally) correct, then passion will determine judgment and people will follow their feelings, even when they lack reasons to support those feelings. (Haidt et al. 2000, p. 2, emphasis added)

Interestingly, a team of psychologists and consequentialist philosophers did attempt to apply Chagnon's strategy to Haidt's findings, and thus to save the consequentialist interpretation of disgust. Stanley, Yin and Sinnott-Armstrong (2019) aim to at least partially rationalize judgments about Haidt's vignettes by also testing how high participants judge the probability of harm to be. In other words, they use a subjectivist consequentialist framework.

> [P]utatively harmless taboo violations are judged to be morally wrong because of the high perceived likelihood that the agents could have caused harm, even though they did not cause harm in actuality. [...] Critically, a manipulation drawing attention to harms that could have occurred (but

[14] For a comprehensive philosophical critique, see Jacobson (2012).

> did not actually occur) systematically increased the severity of moral wrongness judgments. (2019, p. 1)

This debate can be summarized as follows. Haidt notices that disgust is not forward-looking. Since he implicitly holds that only actions from a forward-looking motive are rational actions, he classifies disgust as irrational. Stanley et al. share Haidt's premise that only actions from a forward-looking motive can be rational. They, however, draw the opposite conclusion, viz. to look for a reinterpretation of disgust as forward-looking, which they find in the motive of heuristically justified fear.

Both research teams, I submit, misinterpret the observed behavior. As in Chagnon's case, we have the premise that any rational action must produce something, which then leads to the question what disgust produces, which Haidt et al. seem to answer with "nothing" and Stanley et al. seem to answer with "avoidance of a perceived danger." Once we drop this premise, the puzzle vanishes. The participants' reaction of disgust to, e.g., cannibalism can be understood for what it probably is: taking the fact that *X* is a human being as a reason not to eat *X*—and by so doing, making a noninstrumental rational choice.

Disgust is a difficult philosophical topic, and I freely admit not to possess a theory of it. In table 1, I classify disgust as an interpretative motive, and my *ad hoc* suggestion is to regard it as structurally similar to the other items in that category. The fact that Jack is your friend is a perfectly good reason to buy him a drink; the fact that he is one of us is a perfectly good reason to support his strike, and the fact that he is human is a perfectly good reason not to eat him. All of these facts are proper "supporting reasons," as Haidt calls them, not in need of support from further, instrumental reasons.

I have analyzed an anthropological application of rational choice theory to a backward-looking motive and a psychological application of consequentialism to an interpretative motive. The point of these examples was to support my fourth claim (from section 3), which says that instrumentalist frameworks cause the social sciences to misinter-

pret noninstrumental rational actions. Such misinterpretations are numerous, and some—including those discussed—are influential.

Indirectly, the examples also illustrate what is wrong with their underlying philosophical frameworks. My examples were of detailed and careful applications of instrumentalism to important and universal human motives for actions: revenge and disgust. Their failure illustrates how and why instrumentalism fails as a conception of rational actions. It simply is too narrow. Motives such as justice, kinship and respect for the dead appear to be justifications for actions in their own right. The empirical study of human behavior hence needs to proceed from a paradigm of practical rationality broad enough to accommodate these.

# 7 Objections

No doubt, rational choice theorists and consequentialists will find much to disagree with in my argument. In this last section, I consider three standard objections.

The first objection a consequentialist may raise says that rationality is a criterion of actions rather than their motivations. "Indirect consequentialists" argue "that a consequentialist moral agent need not aim at maximizing the good, nor need be motivated directly by the desire to maximize the good" (Cocking and Oakley, 1995, p. 87).[15] Take the example of vendetta. Why not say that the agent indeed kills *X* because *X* previously killed the agent's brother, i.e., the agent has a backward-looking motive, but that the population-level consequence of their action is to deter potential future killers, i.e. general prevention, and that it is this forward-looking, population-level consequence which makes the action rational?

[15]See, e.g., Adams (1976), Brink (1986), and Goldstick (2002). The *locus classicus* is Sidgwick (1907, section 4.1.1): "[T]he doctrine that Universal Happiness is the ultimate standard must not be understood to imply that Universal Benevolence is the only right or always the best motive of action. For, as we have observed, it is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim."

Various criticisms have been levelled against this understanding of practical rationality, from doubts about the implied empirical claims, such as that punishment deters (Robinson and Darley, 2004), to the accusation that indirect consequentialism requires an "alienated" mindset of the agent (Jollimore, 2003; Williams, 1973, sect. 6) or that it is "self-effacing" (Stocker, 1976). A much more direct reply is available here, however: The discussed problems with instrumentalizing (section 4) reoccur at the population level. If action $A$ is rational because it deters, then $A$ is not a case of revenge. If $A$ is rational because it secures future benefits, then $A$ is not a case of thanking. And if $A$ is rational because those who benefit might aid the agent (or others) in the future, then $A$ is not a display of solidarity. Whether we locate rationality at the level of the individual agent or the population, this point does not change. Hence, if my arguments in section 4 hold at all, then they also hold at the population level.

A second objection, sometimes heard from rational choice theorists, says that their theory is a pure mathematical model and their concept of a preference so minimal as to be almost empty. This objection is more common in economics than in philosophy, and it is usually an implicit assumption rather than an explicitly defended contention, but it is so widespread as to be worth addressing.[16]

It is hard to pin down the exact target of this objection. I take it to object to an understanding of the formalism as either a model of actual or ideal practical choices and an understanding of preferences as egoistic and subjective desires. The latter point does not apply here. I distinguished subjectivism from instrumentalism, and my own concept of an end is so minimal as to include everything from Hume's "passions" to impersonal best states of the the world (see section 2). The other point strikes me as a mere conceptual confusion. A model, by definition, is a model of something. Now, what would that something be if not practical choices? One can, of course, propose a pure formalism that does not model anything in the empirical world, such

[16]For some discussion, see Maialeh (2019) and Angner (2014).

as when, in pure logic or mathematics, an author specifies a set of abstract objects, their interrelations and transformation rules. Qua stipulation, however, such a formalism has nothing whatsoever to do with either rationality or choice and has no application in a social science.

A third objection, advanced by an innocent bystander, might be to regard rational choice theory and consequentialism as providing a partial but in itself correct picture of rational action. After all, no one would deny that instrumental/forward-looking reasoning is a central component of practical rationality. A person who cannot choose adequate means to adequate ends cannot possibly count as a rational agent, even if we concede that this ability is a necessary rather than a sufficient condition. Perhaps one or both paradigms correctly describe the exercise of this ability?

I doubt that proponents of either paradigm would find this stance attractive. Consequentialists propose a general moral theory and, e.g., economists who use rational choice theory regard it as a general framework for economic modelling. The objection essentially suggests to subsume one or both under a third paradigm. It is impossible to judge the plausibility of this without spelling out the details, but it seems obvious that this would amount not to a slight modification of either theory but to the creation of a new theory.

# 8 Conclusion

My starting point was a surprising gap in the literature. Even though the instrumentalist picture of rationality has been massively criticized in metaethics over the past seventy years, no one has analyzed what theories presuppose this picture and are hence affected by these criticisms. The reason for this, I suggested, is that the critics often conflate subjectivism and instrumentalism about practical rationality.

I hope to have shown (i) that any theory which defines rational actions as actions that maximize some outcome is instrumentalist, (ii) that consequentialism and rational choice theory are such theories and

(iii) that there are noninstrumental rational actions which (iv) these two theories and their applications in the social sciences misinterpret. In other words, the discussed problem for instrumental rationality affects numerous theories in philosophy and the social sciences.

## References


Adams, Robert (1976). "Motive utilitarianism." In: *The Journal of Philosophy* 73.14, pp. 467–481.

Albert, Bruce (1989). "Yanomami 'violence': Inclusive fitness or ethnographer's representation?" In: *Current Anthropology* 30.5, pp. 637–640.

Angner, Erik (2014). "To navigate safely in the vast sea of empirical facts." In: *Synthese* 192.11, pp. 3557–3575.

Anscombe, Elizabeth (1981a). "Intention." In: *The Collected Philosophical Papers of G. E. M. Anscombe*. Vol. 2. Oxford: Blackwell, pp. 75–82. [1]1957].

— (1981b). "Modern moral philosophy." In: *The Collected Philosophical Papers of G. E. M. Anscombe*. Vol. 3. Oxford: Blackwell, pp. 26–42. [1]1958].

— (2000). *Intention*. 2nd ed. Cambridge, Mass.: Harvard University Press. [1]1957].

Beccaria, Cesare (1995). "On crimes and punishments." In: *"On Crimes and Punishments" and Other Writings*. Ed. by Richard Bellamy. Trans. by Richard Davies. Cambridge: Cambridge University Press, pp. 1–114. [1]1764].

Bentham, Jeremy (1996). *An Introduction to the Principles of Morals and Legislation*. The Collected Works of Jeremy Bentham, Vol. 1. Ed. by James Burns and Herbert Hart. With an intro. by Fred Rosen. Oxford: Clarendon Press. [1]1789].

Briggs, Rachael (2019). "Normative theories of rational choice: Expected utility." In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta. Fall 2019. [1]2014].

Brink, David (1986). "Utilitarian morality and the personal point of view." In: *The Journal of Philosophy* 83.8, pp. 417–438.

Chagnon, Napoleon (1988). "Life histories, blood revenge, and warfare in a tribal population." In: *Science* 239.4843, pp. 985–992.

— (2012). *The Yanomamö*. 6th ed. Belmont, Cal.: Wadsworth. [1]1968].

Cocking, Dean and Justin Oakley (1995). "Indirect consequentialism, friendship, and the problem of alienation." In: *Ethics* 106.1, pp. 86–111.

Cohon, Rachel (2008). *Hume's Morality. Feeling and Fabrication*. Oxford: Oxford University Press.

Ellis, Anthony (2003). "A deterrence theory of punishment." In: *The Philosophical Quarterly* 53.212, pp. 337–351.

Elster, Jon (1990). "Norms of revenge." In: *Ethics* 100.4, pp. 862–885.

Ferguson, Brian (2001). "Materialist, cultural and biological theories on why Yanomami make war." In: *Anthropological Theory* 1.1, pp. 99–116.

Firth, Raymond (2011). *Primitive Economics of the New Zealand Maori*. London: Routledge. [1]1929].

Foot, Philippa (2001). *Natural Goodness*. Oxford: Clarendon Press.

Gauthier, David (1992). "Artificial virtues and the sensible knave." In: *Hume Studies* 18.2, pp. 401–427.

Goldstick, Daniel (2002). "The 'two hats' problem in consequentialist ethics." In: *Utilitas* 14.1, pp. 108–112.

Goleman, Daniel (1998). *Working with Emotional Intelligence*. Bantam Books.

Haidt, Jonathan (2001). "The emotional dog and its rational tail: A social intuitionist approach to moral judgment." In: *Psychological Review* 108.4, pp. 814–834.

— (2012). *The Righteous Mind. Why Good People Are Divided by Politics and Religion*. New York: Pantheon Books.

Haidt, Jonathan, Fredrik Björklund, and Scott Murphy (2000). "Moral dumbfounding: When intuition finds no reason." In: *Lund Psychological Reports* 1.2, pp. 1–29.

Haidt, Jonathan and Matthew Hersh (2001). "Sexual morality: The cultures and emotions of conservatives and liberals." In: *Journal of Applied Social Psychology* 31.1, pp. 191–221.

Harris, John (1975). "The survival lottery." In: *Philosophy* 50.191, pp. 81–87.

Hume, David (2000). *A Treatise of Human Nature. Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. Ed. by David and Mary Norton. Oxford: Oxford University Press. [1]1739].

Jacobson, Daniel (2012). "Moral dumbfounding and moral stupefaction." In: *Oxford Studies in Normative Ethics*. Ed. by Mark Timmons. Vol. 2. Oxford: Oxford University Press, pp. 289–315.

Jollimore, Troy (2003). "Goldstick on the 'two hats' problem." In: *Utilitas* 15.3, pp. 369–373.

Korsgaard, Christine (1986). "Skepticism about practical reason." In: *The Journal of Philosophy* 83.1, pp. 5–25.

Maialeh, Robin (2019). "Generalization of results and neoclassical rationality: Unresolved controversies of behavioural economics methodology." In: *Quality & Quantity* 53.4, pp. 1743–1761.

Malinowski, Bronislaw (2002). *Collected Works*. Vol. 2: *Crime and Custom in Savage Society*. London: Routledge. [1926].

Mauss, Marcel (2016). *The Gift*. Trans. and annot. by Jane Guyer. With a forew. by Bill Maurer. Expanded edition. Chicago, Ill.: HAU. [1925].

McCloskey, Henry (1957). "An examination of restricted utilitarianism." In: *The Philosophical Review* 66.4, pp. 466–485.

Mill, John Stuart (1985). "Utilitarianism." In: *The Collected Works of John Stuart Mill*. Ed. by John Robson. Vol. 10. Toronto: Toronto University Press, pp. 203–260. [1861].

Nagel, Thomas (1978). *The Possibility of Altruism*. 2nd ed. Princeton, N.J.: Princeton University Press. [1970].

Quinn, Warren (1993). "Putting rationality in its place." In: *Morality and Action*. Ed. by Philippa Foot. Cambridge: Cambridge University Press, pp. 228–255.

Radcliffe, Elizabeth (2018). *Hume, Passion, and Action*. Oxford: Oxford University Press.

Robinson, Paul and John Darley (2004). "Does criminal law deter? A behavioural science investigation." In: *Oxford Journal of Legal Studies* 24.2, pp. 173–205.

Ross, Don (2019). "Game theory." In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta. Winter 2019. [1997].

Sahlins, Marshall (2017). *Stone Age Economics*. With a forew. by David Graeber. 2nd ed. Milton Park, Abingdon, Oxon: Routledge. [1972].

Sidgwick, Henry (1907). *The Method of Ethics*. 7th ed. London: Macmillan. [1874].

Singer, Peter (1972). "Famine, affluence, and morality." In: *Philosophy and Public Affairs* 1.1, pp. 229–243.

Sinnott-Armstrong, Walter (2019). "Consequentialism." In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta. Summer 2019. [2003].

Stanley, Matthew, Siyuan Yin, and Walter Sinnott-Armstrong (2019). "A reason-based explanation for moral dumbfounding." In: *Judgment and Decision Making* 14.2, pp. 120–129.

Stocker, Michael (1976). "The schizophrenia of modern ethical theories." In: *The Journal of Philosophy* 73.14, pp. 453–466.

Tadros, Victor (2011). *The Ends of Harm. The Moral Foundations of Criminal Law*. New York: Oxford University Press.

Teichmann, Roger (2018). "Rational choice theory and backward-looking motives." In: *Virtues and Economics*. Ed. by Peter Róna and László Zsolnai. Cham: Springer, pp. 117–123.

Vogler, Candace (2002). *Reasonably Vicious*. Cambridge, Mass: Harvard University Press.

Weinstein, Eugene (1969). "The development of interpersonal competence." In: *Handbook of Socialization Theory and Research*. Ed. by David Goslin. Chicago: Rand McNally College Publishing, pp. 753–775.

Williams, Bernard (1973). "A critique of utilitarianism." In: *Utilitarianism. For and Against*. Ed. by John Smart and Bernard Williams. Cambridge: Cambridge University Press, pp. 75–150.