

This is a section of [doi:10.7551/mitpress/11252.001.0001](https://doi.org/10.7551/mitpress/11252.001.0001)

# The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

## Citation:

*The Handbook of Rationality*

Edited by: Markus Knauff, Wolfgang Spohn

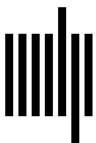
DOI: 10.7551/mitpress/11252.001.0001

ISBN (electronic): 9780262366175

Publisher: The MIT Press

Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.



The MIT Press

## 2.2 Practical and Theoretical Rationality

Ralph Wedgwood

### Summary

Philosophers have long distinguished between practical and theoretical rationality. The first section of this chapter begins by discussing the ways in which this distinction was drawn by Aristotle and Kant; then it sketches what seems to be the general consensus today about how, at least roughly, the distinction should be drawn. The rest of this chapter explores what practical and theoretical rationality have in common: in the second section, several parallels between practical and theoretical rationality are outlined, and it is argued that these parallels make it plausible that a unifying account of rationality can be given. Finally, in the third section, a number of such unifying accounts of practical and theoretical rationality are surveyed. These include accounts that are inspired in various ways by Hume and by the results of formal decision theory, as well as views that appeal to reasons and to the distinctive value of correct or appropriate attitudes.

### 1. The Distinction between Practical and Theoretical Rationality

#### 1.1 Aristotle and Kant

The terms “practical” and “theoretical” derived from ancient Greek philosophy, particularly from the work of Aristotle, who in his ethical writings gave a theory of both *theōria* and *praxis*.

In the relevant senses,<sup>1</sup> both *theōria* and *praxis* are “activities” (*Nicomachean Ethics*, 1146b33) of “the parts of the soul that involve reason” (1095a10). In Aristotle’s teleological system, the nature of both these activities is explained in terms of their essential *function*—that is, the role that these activities *ideally should* play in a human life. Specifically, when it fulfills its essential function, *theōria*—often translated as “contemplation” or “study”—is the manifestation of the virtue of *theoretical wisdom* (*sophia*, 1177a26), which in turn is explained as consisting of *scientific knowledge* (*epistēmē*), resting

on *comprehension* (*nous*) of the first principles of science (1141a18). *Theōria*, then, involves actively exercising one’s understanding of a body of scientific knowledge that one possesses (1177a26).

However, theoretical wisdom (*sophia*) is only *one* of the so-called intellectual virtues. The intellectual virtues also include technical skill (*technē*, 1140a1–25) and practical wisdom or prudence (*phronēsis*, 1140a26–b29). While theoretical wisdom is manifested in contemplation (*theōria*), practical wisdom is manifested in *action* (*praxis*). At least when the agent’s capacities are all fulfilling their essential functions, a *praxis* is the execution of a “decision” (*prohairesis*, 1094a5). According to Aristotle, a “decision” is a special kind of desire to perform an action here and now, which arises from a wish (*boulēsis*, 1111b26) for an end that is viewed as good, through deliberation or rational calculation (*logismos*) about how best to achieve that end (1112b15). When we manifest the virtue of practical wisdom, we make a correct decision, on the basis of grasping the truth about the good and correctly perceiving our practical situation, and by executing such a decision, we perform a correct action (1140b21).

This is the general picture of Aristotle’s distinction between theoretical and practical wisdom, as well as the activities, *theōria* and *praxis*, that under favorable circumstances are their manifestations. However, at one point, he seems to give a rather different general characterization of the difference between theoretical and practical reason (1139a3–15):

Let us assume there are two parts [of the soul] that have reason: with one we study beings whose principles do not admit of being otherwise than they are, and with the other we study beings whose principles admit of being otherwise. . . . Let us call one of these the scientific part, and the other the rationally calculating part; for deliberating is the same as rationally calculating, and no one deliberates about what cannot be otherwise. Hence the rationally calculating part is one part of the part of the soul that has reason.<sup>2</sup>

Here Aristotle seems to say that theoretical reason is concerned with necessary truths, and practical reason—the

sort of reason that consists in the capacity for “deliberating”—is concerned with contingent truths. This characterization strikes contemporary philosophers as puzzling—since it leaves no room for theoretical reasoning about contingent facts that have no practical relevance (such as the contingent facts of ancient history, paleontology, and the like).

In a similar way, Kant also seems to seek an account of the difference between practical and theoretical reason in terms of the “cognitions” (*Erkenntnisse*) that they are concerned with. One short statement of this distinction appears in the presentation of Kant’s *Logic* that was published toward the end of Kant’s life by his colleague Gottlob Jäsche. In the primary sense, according to Kant (1900ff., volume 9, p. 86), a “practical cognition” is an “imperative”—where an imperative is defined as “a proposition that expresses a possible free action, whereby a certain end is to be made real.” By contrast, theoretical cognitions are characterized as follows:

Theoretical cognitions are ones that express not what ought to be but rather what is, hence they have as their object not an acting (*kein Handeln*) but rather a being (*ein Sein*).<sup>3</sup>

Kant (1900ff., volume 4, p. 413) gives a different account of “imperatives” in his *Groundwork of the Metaphysics of Morals*, where he says that an imperative is the “formula” of “a representation of an objective principle, so far as it is necessitating for a will.” What this seems to mean is that an imperative is the expression of a cognition that represents a principle that the will of every rational being ought to follow—or, more precisely, the application of such a principle to an agent who does not necessarily follow or conform to this principle. As he says, “All imperatives are expressed through an *ought*. . . . They say that it would be good to do or refrain from something” (p. 413).

This approach, however, brought Kant into difficulties with his account of so-called *hypothetical* imperatives (p. 414). Every hypothetical imperative presupposes a certain *end* or *purpose* and merely indicates the *means* that are necessary for achieving that end or purpose. In other words, the imperative asserts that a certain action is *good* for the relevant end. Like all imperatives, each hypothetical imperative represents a principle that should in a sense guide the will of all rational beings—at least in the weak sense that no rational beings may will the end without also willing the necessary means.

However, as Kant later pointed out in the “First Introduction” to the *Critique of the Power of Judgment*, each of these hypothetical imperatives is in a way nothing more than a *theoretical* proposition about the causal

connection between an action and the relevant end. His puzzling conclusion is that hypothetical imperatives—which he now prefers to call “technical” rather than “hypothetical”—are practical “in form” but not “in content” (Kant, 1900ff., volume 20, pp. 196, 200).

These difficulties seem to reveal that it is a mistake to account for the difference between practical and theoretical reason in terms of two kinds of propositions (or “cognitions”). Instead, the difference should be explained by distinguishing between two kinds of *attitudes* that can be taken toward propositions. This is the approach that has become the consensus among contemporary philosophers.

## 1.2 Belief versus Practical Attitudes

Virtually all contemporary philosophers seem to agree on the following way of distinguishing between theoretical and practical rationality:

- (a) Theoretical rationality is the rationality of *beliefs* and other belief-like or belief-involving phenomena (such as credences, inferences, and the like).
- (b) Practical rationality is the rationality of the *practical attitudes*—like intentions, decisions (or choices), and preferences—that are directly executed in action.

According to this approach, theoretical and practical rationality do not differ with respect to the *propositions* that they are concerned with. They differ with respect to the kinds of *attitude* that are in question—(a) belief in the case of theoretical rationality and (b) intention or decision or preference in the case of practical rationality.

In appealing to this contrast between (a) belief and (b) intention or decision, are contemporary philosophers departing from the philosophical tradition? There are reasons for thinking that they are not. As we have seen, Aristotle firmly associates theoretical wisdom with *knowledge*, and many contemporary philosophers think of knowledge as in a sense the *best case* of a certain central kind of belief—specifically, of the kind of belief that is sometimes called “full belief.” On the practical side, Aristotle clearly associates practical wisdom with decision and action. Similarly, Kant (1900ff., volume 4, p. 412) states that “the will is nothing other than practical reason”—and what Kant calls a “determination of the will” seems essentially to involve an intention to act.

Still, on both the theoretical and the practical side, several different kinds of attitudes could be considered. On the theoretical side, there is not just full or outright belief but also various levels of confidence or credence that we can take toward the propositions that concern us. In addition, there may be various further belief-like

attitudes that should be considered. Some philosophers—like Joyce (2010)—have suggested that there are both precise and imprecise credences; some—like Edgington (1995)—have argued that there are both conditional and unconditional credences; and others—like Friedman (2013)—have argued that suspension of judgment is a distinctive belief-like attitude as well. If all these different belief-like attitudes exist, they can all be assessed as rational or irrational. If any attitudes of these kinds are rational, they would count as theoretically rational.

In addition to these belief-like mental states, there are also belief-involving *events* of belief revision or theoretical reasoning. Such events involve *changing* one's beliefs in some way—either by coming to have a belief-like attitude toward a proposition toward which one previously had no attitude at all, by replacing an old belief-like attitude that one had in a proposition with a new attitude toward the very same proposition, or simply by reaffirming the old attitude that one had toward that proposition. If any of these mental events are rational, they would also count as theoretically rational.

On the practical side, intentions are mental *states*—while choices or decisions are mental *events* in which one forms or revises one's intentions about how to act. If an intention is executed or carried out, the execution of the intention is an action. It seems that all of these states and events—intentions, decisions (or choices), and actions—can be assessed as rational or irrational. If rational, they would count as practically rational. Some philosophers have argued that, among these states and events, some are more fundamental to practical rationality than others. For example, Harman (1986, p. 77) has argued that practical rationality fundamentally concerns processes in which agents change their plans or intentions, and Broome (2013, p. 250) has argued that requirements of rationality fundamentally apply to sets of mental states, like beliefs and intentions, rather than applying directly to actions.

In formal accounts of practical rationality—such as those of Jeffrey (1983) and Joyce (1999)—the focus is typically not on intentions, decisions, or actions but on *preferences*. In these accounts, the fundamental instances of practical rationality are preferences and events in which our preferences change in response to changes in our beliefs. This approach is also akin to the contemporary consensus, since theorists who focus on preferences typically assume that there is an intimate connection between preferences, on the one hand, and choices (or decisions) and actions, on the other. Specifically, these theorists typically assume that no agent ever chooses (or decides on), or intentionally takes, any

course of action *A* if there is an available alternative *B* that the agent prefers over *A*.

Admittedly, some contemporary philosophers would have reason to reject this way of distinguishing practical and theoretical rationality. Specifically, some philosophers—such as Setiya (2007, p. 49)—actually identify intentions with a special kind of belief. On this view, for you to intend to *F* is just for you to have a belief of a certain kind that you will *F*. What is special about intention, on this view, is that it is a belief that tends in a certain distinctive way to *cause* the truth of its propositional content (unlike most beliefs, which are caused by the truth of their content or else causally independent of the truth of their content). On this view, it could happen that when you have a belief of this sort, to the effect that you will *F*, this belief is *theoretically* rational (if when you have the belief, it is clear to you, given your evidence, that you will indeed *F*), even if it is not *practically* rational (if *F*ing is obviously a foolish or inadvisable course of action). For this reason, these theorists would have to find a different way to distinguish between practical and theoretical rationality.

Otherwise, however, virtually all contemporary theorists would distinguish between practical and theoretical rationality in the way that I have described. According to this contemporary consensus, then, theoretical rationality is the rationality of beliefs, while practical rationality is the rationality of intentions and choices and the like.

## 2. Unifying Practical and Theoretical Rationality?

Practical and theoretical rationality have some striking features in common. First, similar vocabulary is used on both the practical and the theoretical sides. The term “rational” itself is applicable, in what appears to be the very same sense, to both beliefs and choices. We can talk about the beliefs and choices that are “rationally required” of a thinker in a given situation and of the beliefs and choices that are “rationally permissible” for the thinker in the situation. We can talk about “reasonable” or “justified” beliefs and choices—where there is one way of using terms like “reasonable” and “justified” on which they express the very same concept as the term “rational.” Finally, we can talk about “reasons” for beliefs and “reasons” for choices and actions; on both the practical and theoretical sides, we distinguish between there being “some reason” for a belief or an action, “sufficient reason” for the belief or action, and “compelling” or “overriding reason” for the belief or action. It seems unlikely that these terms—“justified,” “reasonable,” and “reason”—are used in different senses when applied to belief as opposed to choice or action.

Moreover, both theoretical and practical rationality come in *degrees*: while some beliefs are only slightly irrational, other beliefs are extremely irrational—and the same is obviously true of choices or decisions as well. On both the theoretical and the practical side, there is some plausibility in the following definition of a “rational requirement”: meeting a certain condition *C* is rationally required of you at a time *t* if and only if, in all relevantly available worlds in which you are thinking as rationally as you can at *t*, you meet condition *C*.<sup>4</sup>

Finally, on both the practical and the theoretical side, we can draw the same distinction between your having beliefs (or intentions) that *it is rational* for you to have and your having those beliefs (or intentions) *rationally*. An irrational agent might, through sheer dumb luck, choose an option that it is rational for him to choose, but he would not thereby be choosing this option rationally. In epistemology, this distinction is typically drawn by means of the terminology of “propositional” and “doxastic” justification: an irrational thinker might, through sheer dumb luck, believe a proposition that there is *propositional* justification for her to believe, but this irrational thinker would not thereby be believing the proposition in a *doxastically* justified manner. Clearly, however, the same distinction can be drawn among choices and intentions as well.

With all these parallels between practical and theoretical rationality, it seems plausible that there should in principle be some way of *unifying* the two domains. What would such a unification of these two domains amount to? First, it would consist of a collection of general principles, capturing features of rationality in general, including both practical and theoretical rationality. Second, it would have to be true that the specific features of theoretical rationality can all be derived by combining these general principles about rationality with some specific principles about belief, while the specific features of practical rationality can all be derived by combining these general principles about rationality with some specific principles about intention, choice, and action. As we shall see in the following section, a number of different attempts have been made to provide a unifying account of this sort.

### 3. Unifying Accounts of Rationality

#### 3.1 Formal Theories of Rationality

Philosophers who have devised formal theories of rationality have typically offered unified accounts of both practical and theoretical rationality. One particularly common idea is that to be rational, an agent’s beliefs

and preferences must be capable of being “*represented*” by a probability function and a utility function. There are some subtle differences between different versions of this idea, but according to the most common version, to say that a probability function and utility function “*represent*” the agent’s beliefs and preferences is to say that the following conditions hold:

- (a) For any two relevant prospects *A* and *B*, the agent prefers *A* over *B* if and only if the utility function assigns a higher utility to *A* than to *B*, and the agent is indifferent between *A* and *B* if and only if the utility function assigns the same utility to *A* as to *B*.
- (b) For any two propositions *p* and *q*, the agent believes *p* with greater confidence than *q* if and only if the probability function assigns a higher probability to *p* than to *q*, and the agent believes *p* and *q* with the same degree of confidence if and only if the probability function assigns the same probability to *p* as to *q*.
- (c) For every prospect *A*, the utility that the utility function assigns to *A* is equal to *A*’s expected utility according to this probability function.

Suppose that every rational agent’s beliefs and preferences can be represented by a utility function and probability function in this way. Then it follows that every rational agent’s preferences give a *complete* ordering of the prospects over which they are defined: that is, for any two such prospects *A* and *B*, the agent prefers *A* over *B*, prefers *B* over *A*, or is indifferent between the two. Rational agents’ preferences must also be *transitive*: if the agent prefers *A* over *B* and either prefers *B* over *C* or is indifferent between *B* and *C*, then the agent must also prefer *A* over *C*.

The centerpiece of many such formal accounts of rationality is a “*representation theorem*”—a formal proof that so long as the agent’s beliefs and preferences meet certain conditions, they can be “*represented*” by a probability function and utility function in something like this sense. One of the most powerful of these theorems is due to Savage (1972). Savage only needs to assume that the agent’s preferences meet a number of conditions—no special assumptions about the agent’s beliefs are required—and on this basis he proves that these preferences can be represented by a *unique* probability function and a utility function that is unique up to an arbitrary choice of a unit and zero point.

How could this account of rationality unify practical and theoretical rationality? One approach might rely on Savage’s theorem and attempt simply to *reduce* theoretical rationality to practical rationality. If it is true that every rational agent’s preferences must satisfy the conditions

that Savage's proof assumes, then a rational agent's beliefs can simply be *identified* with the probability function that is uniquely determined by these preferences. This might suggest a *pragmatist* view of belief—in effect, the view that a practically rational agent has no beliefs except for those that are already implicit in her preferences. If this pragmatist view of belief is correct, and if there is no more to be said about rational belief than that rational beliefs must be capable of being represented by a probability function, then this approach would be able to reduce theoretical rationality to practical rationality in this way.

However, this pragmatist view of belief—the view that the practically rational agent's beliefs are just those that are implicit in her preferences—is intensely controversial. Most more recent proponents of various forms of formal decision theory—such as Joyce (1999, p. 90)—would prefer to avoid being committed to this pragmatist view of belief.

If we reject the pragmatist view of belief, we would presumably accept that there are principles of rational belief that do *not* simply flow from the constraints on rational preferences. So, we would need an alternative way of unifying theoretical and practical rationality. One approach might stay close to the pragmatist tradition by arguing that the requirements of theoretical and practical rationality are both explained by the pragmatic defects of violating those requirements. For example, we might rely on the well-known “money pump” and “Dutch book” arguments, which seek to show that violating any requirements of either practical or theoretical rationality makes us willing to accept a collection of bets that will result in our losing money no matter what happens.<sup>5</sup>

These “Dutch book” and “money pump” arguments are fascinating, but their significance is disputed. Even if your beliefs do make you willing to accept a collection of bets that would result in your losing money whatever happens, you might never be offered such a collection of bets, and your beliefs might provide extremely useful guidance for the practical situations that you are actually in. So, it is controversial whether these arguments can really show what fundamentally unifies practical and theoretical rationality.

### 3.2 Reducing Theoretical Rationality to Instrumental Rationality?

Many contemporary philosophers who work on rationality seek to develop certain key insights of Hume (1739/1740, II.iii.3). The interpretation of Hume is controversial.<sup>6</sup> However, some contemporary philosophers take Hume's key lesson for us to be that the primary

form of rationality is what is often called “instrumental rationality”—the rationality of taking what (at least according to the evidence) seem to be optimal means to one's goals.

If instrumental rationality is primary in this way, then it appears that theoretical rationality must consist in having a set of beliefs that (according to the evidence) seems to be an optimal means to our distinctive *cognitive* or *epistemic* goals. Most versions of this approach assume that the relevant cognitive goal is that of believing the *truth*, and not believing what is *false*, about the questions that we have considered. It is then rationally permissible for a thinker to have a given set of beliefs if and only if these beliefs seem (at least according to the evidence) to be optimal means to this goal.<sup>7</sup>

This approach has also been criticized by a number of philosophers—notably, by Thomas Kelly (2003). Kelly's main concern is that it does not seem true that ordinary thinkers have very general cognitive aims or goals. It is not obviously true that it genuinely is one of my goals simply to believe true propositions as such. My cognitive goals tend to be more local and specific. I have the goal of having a true belief about the times of the flights from Denmark to California—but this is only because I need to make a choice about how to fly from Denmark to California sometime soon. I have no comparable interest in having true beliefs about the times of the flights between Cairo and Moscow, since I currently have no reason to think that I am likely to fly on that route at any point in the foreseeable future.

If Kelly (2003) is right that our cognitive goals are local and specific in this way, then these goals are *idiosyncratic*. Different people will have different cognitive goals. But the requirements of theoretical rationality seem to be in a way *intersubjective*. If you are compelled by an annoying interlocutor to consider a certain proposition *p*, and at the same time you are confronted with compelling evidence for *p*, then it seems that theoretical rationality will require you to believe *p*—even if it was not in any sense one of your goals to have a true belief about *p*. According to Kelly (2003, p. 623), it is hard to see how the instrumentalist conception of theoretical rationality can explain this fact.

For these reasons, then, while the project of reducing theoretical rationality to practical rationality is still worth investigating, few contemporary philosophers are optimistic about its chances of success.

### 3.3 Rationality and Reasons

Many contemporary philosophers seek to understand rationality in terms of *reasons*. Roughly, to think rationally,

on this reasons-centered approach, is to respond appropriately to one's reasons. This approach holds out the hope of unifying practical and theoretical rationality. Practical rationality consists in making choices—and forming, revising, and maintaining one's intentions or plans about how to act—in a way that involves responding appropriately to reasons for such choices and intentions. Theoretical rationality consists in forming, revising, and maintaining one's beliefs in a way that involves responding appropriately to reasons for belief.

A number of variations on this basic theme are possible. Thus, for example, Joseph Raz does not directly identify rationality with responding appropriately to reasons. Instead, Raz (2011, p. 89) identifies *irrationality* with the “malfunctioning” of our “rational powers”—and Raz (2011, p. 85) identifies these “rational powers” with our capacities for recognizing and responding appropriately to reasons. Kolodny (2005, 2008) hopes to reduce rationality to responding correctly to the reasons that one believes there to be.

A different refinement of this approach postulates a distinction between “objective” and “subjective” (or “apparent”) reasons. On one interpretation of this distinction, there is a certain set of propositions—call them the “reason-propositions” for the relevant agent at the time in question—such that the “objective reasons” for the agent at that time can be identified with the reason-propositions that are *true*, and the “subjective reasons” that the agent has at that time can be identified with the reason-propositions that the agent *believes* at that time. Then we could say that rationality consists of responding appropriately to one's subjective (or apparent) reasons.<sup>8</sup>

The attempt to give a unified account of rationality in terms of responding appropriately to reasons has been criticized in many ways. (For some powerful and interesting criticisms, see chapter 2.1 by John Broome, this handbook, and also Broome, 2013, chapters 5 and 6.) Perhaps the most fundamental objection, however, is the following: How exactly does this account of rationality as responding appropriately to (“subjective”) reasons succeed in unifying practical and theoretical rationality? Is the notion of a “reason” itself a sufficiently unified notion?

According to Raz (2011, p. 36), *epistemic* reasons for believing a proposition are so “through being facts that are part of a case for (belief in) its truth.” Even if there are also nonepistemic reasons for belief (for example, if an evil demon credibly threatens to punish me unless I believe something), the epistemic reasons are, in his terminology, “standard reasons” (Raz, 2011, p. 40). By contrast, Raz (1999) says that reasons for *actions* “are facts in virtue of which those actions are good in some respect

and to some degree” (p. 23). In other words, standard reasons for belief have to do with *truth*, while reasons for action have to do with *value*, and Raz (2011, pp. 42–43) denies that truth, or true belief, is a value.

For Raz (1999, pp. 22–23), standard reasons are unified by the role that they play in guiding our thought and by the principle that it is constitutive of being an intelligible thinker and agent in the first place that one has a general disposition to act in ways that one believes to be good, as well as to believe propositions that are supported by truth-related considerations. Raz's ideas on this topic are controversial: some philosophers, like Setiya (2007, pp. 21–67), argue strenuously against Raz's idea that intentional action is always undertaken under “the guise of the good.” So, it is debated whether Raz successfully answers the unification challenge.

Schroeder (2007, p. 29) takes a different approach. He aims to combine the appeal to reasons with the idea that is suggested by Hume's (1739/1740, II.iii.3) focus on *desire*. Thus, he proposes that all reasons are grounded in desires (or in some other similar attitude). Every reason for you to *F* is grounded in some state of affairs *S* that you desire, in such a way that the reason consists of some fact that helps to explain why your *Fing* promotes, or raises the chances of, *S* (compared to the relevant alternative to *Fing*). Schroeder (2007, p. 113) expresses sympathy for the idea that a view of this kind gives a good account of reasons for belief as well as of reasons for action.

On the face of it, this is a surprising account of reasons for belief. But Schroeder ingeniously argues for several auxiliary theses that could explain why this view is more plausible than it first appears. First, a reason for believing *p* only needs to explain why believing *p* raises the chances of some desired state of affairs to some degree (compared to the relevant alternative). So, as Schroeder (2007, p. 113) argues, whatever states of affairs the agent may desire, if a proposition *p* is false, then not believing *p* will at least somewhat raise the chances of these desired states of affairs; so there is always some reason for not believing *p* if *p* is false. Second, he argues that the *weight* of a reason is not automatically determined either by the strength of the desire or by the degree to which the reason raises the chances of the desired state of affairs—instead, he proposes an entirely different account of the weight of reasons (Schroeder, 2007, chapter 7); so even if all reasons for belief are grounded in desires, it may still normally be the case that there is *most reason* for us not to believe *p* if *p* is false.

Schroeder's Hume-inspired account of reasons is still a work in progress. It remains to be seen exactly how an

account developed along these lines can accommodate enough of our intuitive judgments about the rationality of belief, choice, and action. Still, if an account of this kind can be developed in detail, it would be an illuminating way of unifying practical and theoretical rationality.

### 3.4 A Values-Centered Approach

As we saw in section 2, rationality comes in *degrees*. Of the alternative possible beliefs and choices that you might have in response to your cognitive situation, some are more irrational than others. In this sense, rationality “ranks” some beliefs and choices as more rational than others.

This ranking of beliefs and choices will typically play a *reasoning-guiding* role: normally, your judgments of rationality will guide you toward having the beliefs or making the choices that you judge to be more rational and away from the beliefs and choices that you judge to be less rational. In this respect, rationality resembles a *value*, or a way of being good. (Specifically, it resembles the values that rank some alternatives as in the relevant way *better* than others and guide us toward *preferring* the alternatives that we judge to be better over those that we judge to be worse.) Thus, it seems plausible that rationality is itself a value. Rational thinking is thinking that is *good* in a certain way; irrational thinking is thinking that is *bad* in a corresponding way, and in general, the more irrational your thinking is, the *worse* your thinking is in this way.<sup>9</sup>

While we can evaluate both beliefs and choices as rational, there are also other ways of evaluating beliefs and choices. There is one particularly fundamental way of evaluating beliefs and choices that I shall pick out by using the terms “correct” and “incorrect,” or “appropriate” and “inappropriate.” (In everyday language, many other terms are also used: for example, we could talk about your “getting things right” in your belief about a certain question, or about your making “the wrong choice” among a set of options, and so on.)

As I am using the terms here, the key difference between rationality and appropriateness (or correctness) is this: whether or not a belief or choice is *rational* depends solely on its relations to the mental states and events that are present in the thinker’s mind at the relevant times—whereas whether or not a belief or choice is *appropriate* (or correct) depends on its relations to the external world. Thus, whether a *belief* is correct or appropriate depends on whether the proposition that is believed really is true, and whether a *choice* is correct or appropriate depends on whether the chosen option is feasible and really a good thing to do.

Appropriateness also seems to come in degrees: some choices are more inappropriate—more badly wrong or incorrect—than others. I propose that something similar is true of beliefs. Specifically, I propose that what formal epistemologists like Joyce (1998) refer to as a belief’s “degree of inaccuracy” can be interpreted as a measure of the degree to which a belief is incorrect or inappropriate. Thus, if the proposition *p* is true, then the greater the confidence with which you believe *p*, the more appropriate or correct your belief is, while if *p* is false, then the greater the confidence with which you believe it, the more inappropriate or incorrect your belief is. (So, on this interpretation of what it is to “get things right” or to “believe appropriately,” it is possible for an irrational belief to get things right, through sheer dumb luck.)

In reasoning, we can form *expectations* of how correct different alternative beliefs or choices would be, and these expectations can guide us toward the beliefs and choices that we expect to be more correct and away from those that we expect to be less correct. In this way, it seems that the concept of correct or appropriate attitudes is—like the concept of rational attitudes—an alternative-ranking reasoning-guiding concept. This gives us reason to think that it is also a value concept.

It may be plausible that correctness is, to put it metaphorically, the “aim” of rationality. Suppose that for every agent and every point in time, there is a special probability function *P* such that the credences the agent has at that time are rational precisely in proportion to how closely they *approximate* that probability function *P*. Let us call this the “rational probability function” for the agent at that time. Then the key connection between rationality and correctness could be this: irrationality is, according to this rational probability function, *bad news* about correctness. The more irrational your beliefs or choices are, the greater the degree to which this probability function expects your beliefs and choices to fall short of the highest expected degree of correctness that is available to you in your situation.

Some philosophers will certainly object to this approach. As we have seen, Raz (2011, p. 40) denies that true belief is a value, and it may seem strange to say that correctness or appropriateness is a value. It must be conceded that it is a distinctive kind of value: the value of being admirable is exemplified by the *objects* of admiration, and when we judge that an object is admirable, that judgment guides us toward having certain attitudes toward the object in question. By contrast, the value of correctness or appropriateness is exemplified by various different *attitudes* (including attitudes of admiration).



A judgment about which attitudes are correct does not guide us toward having any (higher-order) attitude toward those attitudes; it directly guides us toward having the lower-order attitudes themselves. On reflection, it is clear that judgments about rationality guide us in the same way. Once it is conceded that both correctness and rationality are distinctive in this way, it seems more plausible that correctness and rationality are values, which are connected to each other in the probabilistic way that I have described.

The important point for our present purposes is that this approach to rationality clearly unifies both practical and theoretical rationality: it applies the same account—the idea of irrationality as bad news about correctness—to both practical and theoretical rationality. The difference between practical and theoretical rationality lies in the different standards of correctness that apply to belief and choice: the appropriate object of *belief* is what is *true*, and the appropriate object of *choice* is what is *feasible and good*. The specific features of rational *belief* result from applying this unified account of rationality to the standard of correctness for belief, while the specific features of rational *choice* result from applying the same account of rationality to the standard of correctness for choice. If this account can be successfully defended, then on both the practical and the theoretical side, rationality consists in thinking in a way that provides good news about its own correctness.

### Notes

1. Aristotle seems sometimes to use “*praxis*” in a broader sense in which it includes all voluntary behavior (1111a26). But in this broader sense, action does not require “reason” (since even nonhuman animals and children are capable of *praxis* in this sense). So I shall focus on the narrower usage of the term here.
2. This translation is from Aristotle (1999).
3. These translations are from Kant (1992, 2000, 2012).
4. For a different conception of “rational requirements,” see Broome (2013, chapter 7).
5. For an illuminating discussion of the “Dutch book” and related arguments, see Hájek (2005).
6. Millgram (1995) has powerfully argued that Hume himself was a *nihilist* about practical rationality—that is, he denied that terms like “rational” can strictly and literally be used to evaluate actions and volitions at all.
7. For a clear recent example of this approach, see Steglich-Petersen (2018).

8. One prominent philosopher who gives an account of rationality in terms of what one has sufficient “apparent reason” to do is Parfit (2011, pp. 34–35).

9. For a more detailed account of this values-centered approach, see Wedgwood (2017).

### References

- Aristotle. (1999). *Nicomachean ethics* (T. H. Irwin, Ed., 2nd ed.). Indianapolis, IN: Hackett.
- Broome, J. (2013). *Rationality through reasoning*. Chichester, England: Wiley-Blackwell.
- Edgington, D. (1995). On conditionals. *Mind*, 104(414), 235–329.
- Friedman, J. (2013). Suspended judgment. *Philosophical Studies*, 162(2), 165–181.
- Hájek, A. (2005). Scotching Dutch books? *Philosophical Perspectives*, 19, 140–151.
- Harman, G. (1986). *Change in view*. Cambridge, MA: MIT Press.
- Hume, D. (1739/1740). *Treatise of human nature*. In A. Merivale & P. Millican (Eds.), *Hume texts online*. Retrieved from <https://davidhume.org/>. References are to book, part, and section number.
- Jeffrey, R. C. (1983). *The logic of decision* (2nd ed.). Chicago, IL: University of Chicago Press.
- Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4), 575–603.
- Joyce, J. M. (1999). *Foundations of causal decision theory*. Cambridge, England: Cambridge University Press.
- Joyce, J. M. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives*, 24, 281–323.
- Kant, I. (1900ff.). *Gesammelte Schriften* [Collected writings]. Berlin, Germany: Reimer (later De Gruyter).
- Kant, I. (1992). *Logic* (G. B. Jäsche, Ed.). In J. M. Young (Ed. & Trans.), *Lectures on logic* (pp. 521–642). Cambridge, England: Cambridge University Press.
- Kant, I. (2000). First introduction to the *Critique of the power of judgment*. In P. Guyer (Ed.), *Critique of the power of judgment* (P. Guyer & E. Matthews, Trans., pp. 1–52). Cambridge, England: Cambridge University Press.
- Kant, I. (2012). *Groundwork of the metaphysics of morals* (M. Gregor & J. Timmermann, Eds. & Trans., Rev. ed.). Cambridge, England: Cambridge University Press.
- Kelly, T. (2003). Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research*, 66(3), 612–640.
- Kolodny, N. (2005). Why be rational? *Mind*, 114(455), 509–563.

- Kolodny, N. (2008). Why be disposed to be coherent? *Ethics*, 118(3), 437–463.
- Millgram, E. (1995). Was Hume a Humean? *Hume Studies*, 21(1), 75–94.
- Parfit, D. (2011). *On what matters* (S. Scheffler, Ed., Vol. 1). Oxford, England: Oxford University Press.
- Raz, J. (1999). *Engaging reason*. Oxford, England: Oxford University Press.
- Raz, J. (2011). *From normativity to responsibility*. Oxford, England: Oxford University Press.
- Savage, L. J. (1972). *The foundations of statistics* (2nd rev. ed.). New York, NY: Dover.
- Schroeder, M. (2007). *Slaves of the passions*. Oxford, England: Oxford University Press.
- Setiya, K. (2007). *Reasons without rationalism*. Princeton, NJ: Princeton University Press.
- Steglich-Petersen, A. (2018). Epistemic instrumentalism, permissibility, and reasons for belief. In C. McHugh, J. Way, & D. Whiting (Eds.), *Normativity: epistemic and practical* (pp. 260–280). Oxford, England: Oxford University Press.
- Wedgwood, R. (2017). *The value of rationality*. Oxford, England: Oxford University Press.



© 2021 The Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.

Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.

Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.

Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)

Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.

Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23

LC record available at <https://lcn.loc.gov/2020048455>