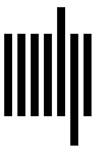
The Handbook of Rationality

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

10.5 Structural Rationality

Julian Nida-Rümelin, Rebecca Gutwald, and Niina Zuber

Summary

This entry focuses on the theory of structural rationality (TSR), which goes beyond the traditional individualist and instrumental approaches to rationality. The starting point of TSR is a shortcoming in traditional rational choice theory (RCT). In RCT, the rationality of an action is evaluated according to its suitability to fulfill the agent's desires at the time of choice, without considering future plans or cooperation. TSR is multidimensional, because it takes at least three dimensions of rational reasoning into account: decisions regarding future aims and plans (diachronic dimension), the plurality of reasons for actions (intrapersonal dimension), and interaction with others (interpersonal dimension). This account of TSR differs from the standard use of "structural rationality" or rather "structural irrationality" in analytic philosophy, since the latter focuses merely on formal relations between different entities such as desires, actions, and intentions. TSR, however, is more radical: it transcends the reductive narrowness of instrumental rationality without denying its practical impact.

1. Individualistic and Nonindividualistic Approaches to Rationality

There are two major traditions in the theory of practical rationality: the first characterizes practical rationality in terms of the acting individual's properties, capabilities, and states, whereas the second starts from social roles, established practices, and institutions. Within both of these traditions, there are numerous variations but few commonalities. This conflict of paradigms ultimately corresponds to a conflict of disciplines: classical and neoclassical economic theory are committed to the individualistic paradigm, while sociology, cultural studies, and social psychology are committed to the second paradigm (cf. chapter 10.4 by Raub, this handbook). Of course, exceptions do exist: the flourishing of behavioral economics (cf. chapter 9.4 by Dhami & al-Nowaihi, this handbook) is a reaction to the deficits of the first paradigm, trying to also incorporate consideration of the second paradigm, for instance, by including nonrational behavior in economic explanation or adherence to social norms (Diamond & Vartiainen, 2007; Gigerenzer & Selten, 2001; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 2000). The theory of structural rationality (TSR) is an account that overcomes this conflict of paradigms. It takes into consideration the embeddedness of individual actions (first paradigm) within the behavioral structures of forms of life (second paradigm). This requires a comprehensive redescription of rational agency.

2. The Paradigmatic Core of the Theory of Structural Rationality

In his *Nicomachean Ethics*, Aristotle (2000) characterizes *akrasia* (which is usually translated as "weakness of the will," although "lack of self-control" would be more accurate) as a singular action that is not embedded in an intended or desired structure of the agent's course of action. In this context, it is essential to take into consideration that Aristotle and the entire classical antiquity as well as medieval and even early modern practical philosophy do not distinguish between a theory of rationality and a theory of morality: living the righteous life of the *dikaios* consists in realizing *eudaimonia* (usually translated as "happiness"). This characterization of a virtuous life, however, is not individualistic, because it depends on the political community (*polis*) as a whole, in which the lives of individuals unfold. Determining what is good for a community is the precondition for determining what is good for the individual. In other words: aspects of structural rationality can already be found in classical accounts of rationality, ethics, and politics (cf. chapter 12.1 by Fehige & Wessels, this handbook).

The breakup of the unity of morality and rationality that originated in the European era of Enlightenment is most radical in the Kantian and the Utilitarian traditions. Jeremy Bentham (1780/2007) argues that pursuing

one's own utility for a maximally advantageous pleasure–pain balance is the decisive motive for individual human action. He deduces the totality of well-being as the guiding moral principle. This tension prevails in contemporary utilitarianism: what is good is determined on the basis of individual well-being or preference fulfillment (Hare, 1963, 1970, 1984; Harsanyi, 1977a, 1977b; Singer, 1993, 2016), while moral obligation does not consist in optimizing individual well-being but in optimizing the *sum* of individual well-being. Rational choice theory (RCT) also focuses on individual preferences: actors do not necessarily optimize their individual well-being but rather strive for the fulfillment of their actual, subjective preferences.

The unity of rationality and morality also disintegrates in the Kantian tradition: rationality amounts to optimizing one's own well-being and is guided by pragmatic imperatives, as Kant (1785) calls them, while moral action is characterized as being exclusively motivated by respect for the Moral Law requiring that the individual maxim of an action is suitable for being transformed into a universal law according to which everybody could act.

In the Kantian tradition, structural embeddedness is realized by adopting only such maxims that are suitable for application as universal rules for action. Kantianism, however, suffers from being underdetermined, because the Moral Law can be met by almost any decision if it is described appropriately. This amounts to the critique of being formalist that Kantianism has been facing since Hegel (1820/2017) and that endures in contemporary communitarianism (Taylor, 1975, 1989). Utilitarianism, in turn, suffers from collectivism, which is incompatible with what John Rawls (1971) called *the separateness of persons*, because it demands maximization of total utility. Individual motives and plans of life are not considered.

It may seem surprising, but ordinary game theory, in tying rational decision making to equilibrium points, also already incorporates an aspect of TSR and is therefore not reducible to decision theory (cf. chapter 9.2 by Perea, this handbook). In Harsanyi's game-theoretic version of rule utilitarianism ("Ethical Bayesianism"; Harsanyi, 1967, 1977c, 1979), the criterion of moral action is the joint optimization of a collective value function through coordinated individual practice. In this understanding, rule utilitarianism is different from act utilitarianism as it requires structurally rational action, that is, to act in such a way that the combination with the actions of others who are also ethically motivated would optimize the common value function (the sum of preference fulfillment), whereas act utilitarianism is about optimizing overall preference fulfillment with every single act.

3. Rational Choice and Structural Rationality

RCT almost never, or at least never directly, includes the embeddedness of pointwise actions in diachronic and interpersonal structures. Other agents are considered part of the environment that influences the fulfillment conditions of personal preferences. However, RCT deals with not only situations of risk and uncertainty but also interaction. Since this branch of RCT emerged historically from the mathematical analysis of games, it is still labeled, misleadingly, as "game theory" (Hargreaves-Heap & Varoufakis, 2004). Rational choice is determined relative to a given value function of consequences and the subjective probability function of the circumstances. In situations of interaction, these circumstances usually consist in the decisions of other agents (possibly plus additional chance events). Hence, the expansion of the rational choice paradigm from decision-theoretic optimization ("playing against nature") to game-theoretic optimization appears to be plausible, at least with regard to interactions in which the decisions of the players are mutually independent from each other (noncooperative games; cf. chapter 9.1 by Albert & Kliemt, this handbook). Problems arise, however, because every agent must assume that the other agents decide rationally, too. The decisions of other players ideally follow the same game-theoretic criteria of rationality. Therefore, game theory renders only those decisions rational that are part of an *equilibrium point*, that is, a combination of decisions in which none of the agents involved has an interest in revoking his decision, as long as all the other agents involved stick to their original decisions (Nida-Rümelin, 1994).

According to game theory, there is only one rational decision for both players, A and B, in the prisoner's dilemma (PD) game described in table 10.5.1: the combination of dominant choices *D*. A vast evidence (Andreoni & Miller, 1993; Barreda-Tarrazona, Jaramillo-Gutiérrez, Pavan, & Sabater-Grande, 2017; Clark & Sefton, 2001; Cooper, Dejong, Forsythe, & Ross, 1996; Kreps, Milgrom, Roberts, & Wilson, 1982; Pothos, Perry, Corr, Matthew, & Busemeyer, 2001), however, shows that a large number of agents chooses *C*, even in one-shot prisoner's

Table 10.5.1
Prisoner's dilemma

		Player B	
		C	*D*
Player A	*C*	3, 3	1, 4
	D	4, 1	2, 2

dilemmas. The orthodox interpretation takes this as an indicator that irrational behavior prevails, but this is simply not convincing. Decision theory and game theory understand themselves as neutral regarding motivation. Criteria of rationality concern choosing the best means to reach one's goals (i.e., to fulfill one's preferences). Imagine a person who wants to cooperate in a one-shot prisoner's dilemma. Asked about her motivation for choosing *C*, she might answer, "I choose *C* because I want that we both choose *C* and because I expect the other person to choose *C*, too." A general theory of rationality should allow for a motivation of this cooperative kind (cf. chapter 10.2 by Schmid, this handbook).

Cooperation is a paradigmatic case of structurally rational agency. It is impossible to characterize cooperation adequately if one takes merely outcomes and probabilities into consideration (Nida-Rümelin, 1993, chapter 4; 2019). To render cooperation (i.e., the choice of dominated *C*) rational in PD, given the cooperative motivations of the players, we cannot dismiss the information of the format itself: the structure of interaction is an essential part of the description of strategies as cooperative or defective. TSR depends on a comprehensive description of strategic options. Cooperative action embeds the individual choice into intended structures of interaction: I do my part, hoping that the other does his part to realize a collective action that secures cooperation (i.e., an outcome that is better for each than the outcome of individual optimization would be).

Assuming a form of "collective" or "shared" intention represents one way of dealing with the PD challenge (Hurley, 2005). Philosophers of collective action and intentionality like Raimo Tuomela (2013), Margaret Gilbert (2013), Michael Bratman (2014), or David Regan (1980; see also Nida-Rümelin, 2014) have therefore introduced a collective perspective on intention, which changes the very outlook on collective-action puzzles like the PD. Collective intention amounts to changing individual utilities based on a group's ethos, as Tuomela has called it.

Empirical findings (Axelrod, 2006; Gintis, Bowles, Boyd, & Richerson, 2005; Rapoport, Chammah, & Orwant, 1956; Seabright, 2004; Sethi & Somanathan, 2005) regarding the behavior of persons in PD situations are best explained by interpreting it as structurally rational. The individual agents prefer *C* to *D* in order to realize $\langle C, C \rangle$. This describes why participants in PD games claim to be disappointed if the other person does not choose *C* and also if they chose *C*.

4. Meta-Preferences

Structurally rational actions challenge traditional RCT, particularly the revealed-preference concept and the belief–desire model of rationality (Nida-Rümelin, 1991). A prominent approach to remedy this problem is the introduction of meta-preferences, for instance, in the works of Amartya Sen and Harry Frankfurt. Sen (1974) argues that there are three levels of morality: the lowest level, namely, the egoist level, which is captured in PD meta-preferences, that is, meta-preferences that accord with first-order preferences. Second, assurance-game (AG) preferences manifest themselves in favoring cooperation to defection if the other person chooses *C*, and third, other-regarding meta-preferences (OR) translate as choosing *C* no matter how the other person decides. Both of these PD variants can be interpreted as disclosing structural rationality: AG players are conditional cooperators, because they do not want to be exploited by egoists, if they cooperate. OR players embed their respective individual choices into the preferred structure of interaction (universal cooperation), possibly with a moral motivation of the Kantian type.

Another prominent account of meta-preferences that is sometimes thought to adequately describe structural aspects of rationality without dismissing the Humean belief–desire model of rationality is presented by Frankfurt (1971). He proposes a concept of a person as an individual who develops meta-preferences, which means that she can distance herself from first-order preferences if she has freedom of the will. Second-order volitions are meta-preferences regarding first-order action-guiding preferences. The idea is that a person is different from a "wanton" insofar as she develops second-order volitions that frame her first-order action-guiding preferences. According to Frankfurt, a "wanton" is not a person in the full sense, since she does not care about her will and only follows her first-order desires.

Both types of meta-preferences display structural aspects of traditional RCT. Without structural reasoning, an individual would be nothing more than a pointwise optimizer and would not be perceived as a person who persists through time and through different moments of interactions (Nida-Rümelin, 2005a, chapter III). We establish and maintain structural deliberation by evaluating our action-guiding preferences. If it turns out that our preferences are structurally incoherent, we are critical about incoherence and wish we had had different first-order action-guiding preferences. This evaluation may influence our first-order action-guiding preferences. If it does not, there is a conflict of the structural dimensions described above and classical rational choice.

5. The Role of Reasons within the Theory of Structural Rationality

The distinction between the theory of rationality and morality collapses if the rationality of an action is characterized in terms of practical reasons (i.e., of different reasons for specific actions). The observation that there are good reasons for actions that are not consequentialist can lead to quite different reactions:

(1) One might assume that these nonconsequentialist reasons are merely *prima facie* reasons that cannot be integrated and thus do not constitute genuine good reasons (this account could be called "consequentialist"). The problem with this account is that some of our most central types of reasons would have to be excluded.

(2) One might confine the range of application of rational choice and exclude moral and other types of good reasons for action (this could be called the "narrow rationality account"). RCT would then not be an all-embracing theory of practical rationality.

(3) One might give up some of the axioms constitutive for standard RCT, as Edward McClennen (1990) demonstrated in his theory of resolute choice (this account could be called "revisionist").

(4) One might redesign the conceptual framework, that is, by reinterpreting the basic concepts of RC such that nonconsequentialist reasons for action can be integrated (this is the strategy of TSR; Nida-Rümelin, 2000, 2005b).

Take the following example: if I have given a promise, I have a reason to keep it even if keeping it has less than optimal consequences for me or the person I gave it to. Speech act theory (Searle, 1969) has analyzed in detail the normative elements that constitute the institution of giving a promise. This analysis reveals the structural character of reasons to act that are connected to the speech act of giving a promise: it maintains the diachronic consistency of one's practice in certain types of interaction. Keeping a promise generates a duty that allows people to relate and forecast the actions of others within a larger context. This institution allows for diachronically and interpersonally structured rationality.

The same applies to communitarian duties one may have because of one's social role, such as parent, teacher, or politician. Communitarian duties constitute these roles and sustain an interpersonally and diachronically consistent social practice. Individual rights can be interpreted structurally, too. They secure the individual's authorship in social contexts. They enable individuals to act within the boundaries of their individual rights. Individual rights thus secure structures of individual agency.

To sum up: the concept of structure is not a kind of external appraisal function that is added to the established life-world practice of giving reasons and accepting reasons. It is a systematization of this practice. We strive for TSR *avant la lettre*. Without a reference to structures, there is no consistent practice, no personhood, and no collective rationality.

6. Is Structural Rationality Instrumental?

The term "structural rationality" can be found in other accounts within the theory of action. The concepts of "structural rationality" in this literature are, however, merely instrumental. Thus, they focus on individual rationality and emphasize the relational and consistent properties of rationality. This instrumental understanding refers to *intra*personal relations between conative attitudes; only some approaches include diachronic dimensions. The analytic philosophical tradition sticks to this account of structural rationality, whereas the TSR also highlights *inter*personal structural features by considering interactions among individuals as a decisive part of rationality. In addition, TSR makes substantial claims about structural rationality.

The planning theory of Michael Bratman (2014) is one of the few accounts that focus on the diachronic and interpersonal aspects of rationality. Bratman thus acknowledges social rationality norms with regard to intentions and policies: friendship, love, dancing together, conversations, getting married, and so on are all practices and forms of action that cannot be explained without reference to a form of "modest sociality" (Bratman, 2007). His concept of modest sociality provides the explanatory basis for our cooperative activities by recourse to conceptual and normative resources that are already in place in our understanding of individual agency. This makes Bratman's explication of cooperation a conceptual extension of his theory of individual agency, since the individual finds himself always already located within social contexts. Bratman argues that collective intentional agency (i.e., cooperative action) is established by shared intentions. A group sharing an intention goes beyond merely intending to do the same thing. They can only act together, or in concert, if they adopt a suitable planning structure that coordinates their actions and deliberations (and if this is common knowledge). If neither of them fulfills the relevant (sub)plan, cooperation fails. Bratman (2007, p. 8), therefore, applies a quite particular notion

of "structures": intentions are considered *plan states* that take up a certain position in coordinating plans that, in turn, establish diachronic and interpersonal structures of agency. Intentions have a coordinating role in managing conduct and providing continuity and organization over time—either in collective action or in individual agency.

Thomas Scanlon (2007) explicitly uses the term "structural irrationality" as an identifier for occurrences of irrational behavior. Scanlon is commonly viewed as having coined the term for analytic philosophy of action.[1] In the tradition of analytic philosophy, the term "structural rationality" is used as a formal, relational concept of consistency, for instance, referring to the consistency of the attitudes, intentions, desires, and so on that an individual holds simultaneously at one particular point in time (Wallace, 2018). Structural rationality does *not* make any substantive claim about the desirability of the action's content; it concentrates on the relational requirements that the agent needs to fulfill if she does not want to act irrationally.

To illustrate this relational form of rationality, take the following example, introduced by D. J. Langlois (2014) as one form of irrationality that he calls "intention inconsistency": your intention is to spend the weekend with your family, while you simultaneously have the intention to complete your new manuscript by Sunday. Also, and this is crucial, you believe that you cannot do both. Irrationality arises due to the relational link between the two intentions and the belief about their incompatibility. Structural irrationality comes into existence because it is implausible to state that one considers such an arrangement as consistent. To accept all these propositions simultaneously is not possible since they exclude each other; they are incompatible with each other. This is, however, not due to their content but due to their relations to each other. Neither of them is irrational per se (i.e., can be criticized by its content); they turn out to be irrational because together they cannot form a balanced structure.

Structural rationality is understood in this manner as a nonsubstantial, instrumental, and (at least in some form) normative requirement of rationality. Other forms of structural irrationality are defined as belief inconsistency or *enkrasia*. These types of irrationality have in common that the irrationality arises due to the incompatible relationships between singular entities, such as intentions, attitudes, and beliefs. Structural irrationality is therefore a problem of instrumental rationality since we cannot criticize the contents of our desires. We can only behave irrationally due to not correctly arranging our conative attitudes, including means–end relations.

John Broome (1999), too, describes the concept of structural irrationality as the failure of some people to consistently combine and arrange their attitudes, which prevents them from structuring their attitudes rationally. Thus, irrationality involves not only forms of refraining from action, although one has good reasons to act (*practical irrationality*), or acting without knowledge or on too little information (*theoretical irrationality*). As with Scanlon's account, irrationality affects not only one or several attitudes but also their arrangement. Such an understanding of structural irrationality excludes consideration of reasons and only pertains to the relations of attitudes. Hence, Scanlon claims that there are certain forms of structural requirements that pertain to the relations between an agent's attitudes. This is why Scanlon mainly speaks about forms of structural *ir*rationality: irrationality is structural if one fails to consider the adequate relations of attitudes.

Highlighting the consistent internal relations of rationality reduces practical reasoning to instrumental reasoning. Hence, one of the most fundamental questions regarding the conceptualization of rationality pertains to the rules that determine how attitudes should be combined and arranged—without referring to their content or their objectives.[2]

7. Is Structural Rationality Ramsey-Compatible?

Modern RCT is based on the utility theorem that was originally proven by the mathematician and philosopher Frank P. Ramsey (1978). John von Neumann and Oskar Morgenstern (1947) revived this concept of rational choice, which was later implemented in different versions by Marschak (1946), Luce and Raiffa (1957), and others. The utility theorem shows that it is possible to transform the qualitative concept of preference into a quantitative concept of utility if conditions of consistency, such as transitivity of preferences, monotonicity, and continuity of preferences regarding lotteries, are met (cf. chapter 8.1 by Grüne-Yanoff and chapter 8.2 by Peterson, both in this handbook). In economic application, this theory is mainly used to express motivational neutrality of utility functions attributed to individual agents (i.e., it is irrelevant how individuals are motivated). It is viewed as sufficient to observe realized preferences. In fact, utility theory requires persons only to have consistent preferences (in the sense of the axioms); it does not, however, require them to have specific motivations. Furthermore, altruistic preferences can be described by these axioms and therefore incorporated in a cardinal, real-valued utility function.[3] The rationality of expected utility maximization is deduced from

the rationality of individual preferences (via the utility theorem), and individual preferences are rational if these preferences are consistent (i.e., meet the axioms of the utility theorem).

TSR is not only instrumental. If expected utility maximization is the criterion of instrumental rationality, then it seems that TSR cannot be Ramsey-compatible. This assumption, however, is fallacious, since it assumes expected utility maximization, as defined by Ramsey, von Neumann, and Morgenstern, to be instrumental. However, preferences can reflect motivations related to the TSR of an individual as well as comply with Rossian *prima facie* duties (Ross, 2002) or with the Kantian categorical imperative or with Aristotelian rules of social practices and still be compatible with the axioms of the utility theorem. Therefore, preferences are compatible with the utility theorem even if they are not instrumental in nature. Hence, expected utility maximization is compatible with TSR. TSR is therefore compatible with rational choice theory. If a structurally rational person makes a decision and the relevant probabilities meet the Kolmogorov axioms and the subjective preferences of the decision maker meet the Ramsey conditions, then this person maximizes expected utility (or, to use a more adequate term: subjective value; cf. chapter 4.1 by Hájek & Staffel, this handbook). Hence, TSR integrates the core idea of RCT: the coherence of two types of subjective attitudes—epistemic (probabilities) and prohairetic (preferences).

8. Structural Rationality—Descriptive or Normative?

There is one salient similarity between traditional RCT and TSR: both are normative and descriptive at the same time. This feature is most conspicuous in economic theory, where economic rationality is used to describe economic behavior and at the same time to advise it (Spohn, 2002).

Looking at the two-sided interpretation of RCT, it comes as no surprise that TSR also qualifies as both normative and descriptive. It starts with our empirically observable behavior and tries to provide an adequate description while simultaneously pointing out and criticizing structurally irrational behavior. It attacks RCT for neglecting structures and therefore disregarding important aspects of any coherent practice. We can see how structures affect behavior in cases of regret or repentance. Structures within theories of rational choice are more successful in describing our life-world reasons for acting. These reasons are, of course, not only empirical but also normative. However, as participants in a specific form of life, we must take these reasons seriously. We can be convinced by arguments that some of these life-world reasons are inadequate or incompatible with other reasons that we might consider to be more fundamental, but we can never leave the normative frame of the life form we participate in (Nida-Rümelin, 2009). The conceptual frame of TSR is better suited to integrate and systematize life-world practical reasons than conventional RCT.

There are duties that relate to social roles, obligations that are related to personal commitments, and still others that originate from prescriptions or rights. The idea that all of these *prima facie* duties could be reduced to the normative criterion of optimizing states of affairs, as in RCT, has a rationalistic turn of the worst kind. Compared to this, TSR is modest. It avoids philosophical and theoretical hypocrisy; it takes practices and persons seriously but, nevertheless, preserves the mathematical core of decision theory (Nida-Rümelin, 1997a). The agent acts rationally if the series of her singular acts constitutes an intended structure (a pattern of actions) that she can accept as part of her *form of life*.[4] Philosophical or economic theory should not prescribe which practical reasons constitute a form of life.

Notes

The main body of the text was written by Julian Nida-Rümelin, whereas Rebecca Gutwald and Niina Zuber surveyed the literature and completed the text, focusing on alternative accounts of structural rationality in the contemporary debate.

1. In fact, this term was introduced much earlier in German philosophy in the context of the critique of consequentialism in ethics and rationality theory (cf. Nida-Rümelin, 1993; accompanied by various articles in the 1990s such as Nida-Rümelin, 1991, 1994, 1997b, 1997c, 1997d). The core elements of structural rationality were later presented in an essay on structural rationality (Nida-Rümelin, 2001; English version: Nida-Rümelin, 2019, Part I). It has been extended to a general theory of practical reason (Nida-Rümelin, 2020).

2. Benjamin Kiesewetter undertakes a thorough analysis of structural irrationality and structural requirements of rationality in his *The Normativity of Rationality* (Kiesewetter, 2017).

3. In fact, infinitely many of those functions can be transformed into one another by positive linear transformations.

4. See also Korsgaard (1996) for an account of autonomy and rationality.

References

Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *Economic Journal*, *103*(418), 570–585.

Aristotle. (2000). *Nicomachean ethics* (W. D. Ross, Trans.). Adelaide, Australia: University of Adelaide Library.

Axelrod, R. (2006). *The evolution of cooperation*. New York, NY: Basic Books.

Barreda-Tarrazona, I., Jaramillo-Gutiérrez, A., Pavan, M., & Sabater-Grande, G. (2017). Individual characteristics vs. experience: An experimental study on cooperation in prisoner's dilemma. *Frontiers in Psychology*, *8*, 1–13.

Bentham, J. (2007). *An introduction to the principles of morals and legislation*. Mineola, NY: Dover. (Original work published 1789)

Bratman, M. E. (2007). *Structures of agency*. Oxford, England: Oxford University Press.

Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford, England: Oxford University Press.

Broome, J. (1999). Normative requirements. *Ratio*, *12*, 398–419.

Clark, K., & Sefton, M. (2001). The sequential prisoner's dilemma: Evidence on reciprocation. *Economic Journal*, *111*(468), 51–68.

Cooper, R., Dejong, D. V., Forsythe, R., & Ross, T. W. (1996). Cooperation without reputation: Experimental evidence from prisoner's dilemma games. *Games and Economic Behavior*, *12*, 187–218.

Diamond, P., & Vartiainen, H. (Eds.). 2007). *Behavioral economics and its applications*. Princeton, NJ: Princeton University Press.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, *68*, 5–20.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox* (Dahlem Workshop Reports). Cambridge, MA: MIT Press.

Gilbert, M. (2013). *Joint commitment: How we make the social world*. Oxford, England: Oxford University Press.

Gintis, H., Bowles, S., Boyd, R., & Richerson, P. J. (2005). The evolution of altruistic punishment. In H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral sentiments and material interests: The foundations of cooperation in economic life* (pp. 215–227). Cambridge, MA: MIT Press.

Hare, R. M. (1963). *Freedom and reason*. Oxford, England: Oxford University Press.

Hare, R. M. (1970). *The language of morals*. Oxford, England: Oxford University Press.

Hare, R. M. (1984). *Moral thinking: Its levels, method, and point*. Oxford, England: Clarendon Press.

Hargreaves-Heap, S. P., & Varoufakis, Y. (2004). *Game theory: A critical introduction*. London, England: Routledge.

Harsanyi, J. C. (1967). Games with incomplete information played by "Bayesian" players. *Management Science*, *14*, 159–182.

Harsanyi, J. C. (1977a). *Morality and the theory of rational behavior*. Oakland: University of California Press.

Harsanyi, J. C. (1977b). *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge, England: Cambridge University Press.

Harsanyi, J. C. (1977c). Rule utilitarianism and decision theory. *Erkenntnis*, *11*(1), 25–53.

Harsanyi, J. C. (1979). *Rule utilitarianism, rights, obligations, and the theory of rational behavior*. Berkeley: Center for Research in Management Science, University of California.

Hegel, G. W. F. (2017). *Grundlinien einer Philosophie des Rechts* [Elements of the philosophy of right]. Hamburg, Germany: Meiner. (Original work published 1820)

Hurley, S. (2005). Rational agency, cooperation, and mindreading. In N. Gold (Ed.), *Teamwork: Multi-disciplinary perspectives* (pp. 200–215). New York, NY: Palgrave Macmillan.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (2000). *Choices, values, and frames*. Cambridge, England: Cambridge University Press.

Kant, I. (1785). *Grundlegung zur Metaphysik der Sitten* [Groundwork of the metaphysics of morals] (Akademie-Ausgabe, Vol. IV, pp. 385–463). Berlin, Germany: Königlich Preußische Akademie der Wissenschaften.

Kiesewetter, B. (2017). *The normativity of rationality*. Oxford, England: Oxford University Press.

Korsgaard, C. (1996). *The sources of normativity*. Cambridge, England: Cambridge University Press.

Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, *27*(2), 245–252.

Langlois, D. J. (2014). *The normativity of structural rationality* (Unpublished doctoral dissertation). Harvard University, Cambridge, MA. Retrieved from http://nrs.harvard.edu/urn-3:HUL.InstRepos:13067678

Luce, R. D., & Raiffa, H. (1957). *Games and decision*. New York, NY: Wiley.

Marschak, J. (1946). Neumann's and Morgenstern's new approach to static economics. *Journal of Political Economy*, *54*(2), 97–115.

McClennen, E. F. (1990). *Rationality and dynamic choice*. Cambridge, England: Cambridge University Press.

Nida-Rümelin, J. (1991). Practical reason or meta-preferences? An undogmatic defense of Kantian morality. *Theory and Decision*, *30*(2), 133–162.

Nida-Rümelin, J. (1993). *Kritik des Konsequentialismus* [Criticizing consequentialism]. Munich, Germany: Oldenbourg.

Nida-Rümelin, J. (1994). Rational choice: Extensions and revisions. *Ratio*, *7*(2), 122–144.

Nida-Rümelin, J. (1997a). *Economic rationality and practical reason*. Dordrecht, Netherlands: Kluwer.

Nida-Rümelin, J. (1997b). Structural rationality, democratic citizenship and the new Europe. In P. B. Lehning & A. Weale (Eds.), *Citizenship, democracy and justice in the new Europe* (pp. 34–49). London, England: Routledge.

Nida-Rümelin, J. (1997c). Structural rationality in game theory. In W. Leinfellner & E. Köhler (Eds.), *Game theory, experience, rationality: Foundations of social sciences, economics and ethics: In honor of John C. Harsanyi* (pp. 81–93). Dordrecht, Netherlands: Kluwer.

Nida-Rümelin, J. (1997d). Why consequentialism fails. In G. Holmström-Hintikka & R. Tuomela (Eds.), *Contemporary action theory: Vol. 2. Social action* (pp. 295–308). Dordrecht, Netherlands: Kluwer.

Nida-Rümelin, J. (2000). Rationality: Coherence and structure. In J. Nida-Rümelin & W. Spohn (Eds.), *Rationality, rules, and structures* (pp. 1–16). Dordrecht, Netherlands: Kluwer.

Nida-Rümelin, J. (2001). *Strukturelle Rationalität* [Structural rationality]. Stuttgart, Germany: Reclam.

Nida-Rümelin, J. (2005a). *Über menschliche Freiheit* [On human freedom]. Stuttgart, Germany: Reclam.

Nida-Rümelin, J. (2005b). Why rational deontological action optimizes subjective value. *ProtoSociology*, *21*, 182–193.

Nida-Rümelin, J. (2009). *Philosophie und Lebensform* [Philosophy and life-form]. Frankfurt/Main, Germany: Suhrkamp.

Nida-Rümelin, J. (2014). Structural rationality and collective intentionality. In S. R. Chant, F. Hindriks, & G. Preyer (Eds.), *From individual to collective intentionality: New essays* (pp. 207–222). Oxford, England: Oxford University Press.

Nida-Rümelin, J. (2019). *Structural rationality and other essays on practical reason* (Theory and Decision Library A: Rational Choice in Practical Philosophy and Philosophy of Science). Cham, Switzerland: Springer.

Nida-Rümelin, J. (2020). *Eine Theorie praktischer Vernunft* [A theory of practical reason]. Berlin, Germany: De Gruyter.

Pothos, E. M., Perry, G., Corr, P. J., Matthew, M. R., & Busemeyer, J. R. (2001). Understanding cooperation in the prisoner's dilemma game. *Personality and Individual Differences*, *51*(3), 210–215.

Ramsey, F. P. (1978). *Foundations: Essays in philosophy, logic, mathematics and economics* (D. H. Mellor, Ed.). London, England: Routledge & Kegan Paul.

Rapoport, A., Chammah, A. M., & Orwant, C. J. (1956). *Prisoner's dilemma: A study in conflict and cooperation*. Ann Arbor: University of Michigan Press.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Regan, D. (1980). *Utilitarianism and co-operation*. Oxford, England: Clarendon Press.

Ross, W. D. (2002). *The right and the good*. Oxford, England: Clarendon Press. (Original work published 1930)

Scanlon, T. M. (2007). Structural irrationality. In G. Brennan, R. Goodin, F. Jackson, & M. Smith (Eds.), *Common minds: Themes from the philosophy of Philip Pettit* (pp. 84–103). Oxford, England: Clarendon Press.

Seabright, P. (2004). *The company of strangers: A natural history of economic life*. Princeton, NJ: Princeton University Press.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, England: Cambridge University Press.

Sen, A. (1974). Choice, orderings and morality. In S. Körner (Ed.), *Practical reason* (pp. 54–67). Oxford, England: Blackwell.

Sethi, R., & Somanathan, E. (2005). Norm compliance and strong reciprocity. In H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral sentiments and material interests: The foundations of cooperation in economic life* (pp. 229–250). Cambridge, MA: MIT Press.

Singer, P. (1993). *Practical ethics*. Cambridge, England: Cambridge University Press.

Singer, P. (2016). *The most good you can do: How effective altruism is changing ideas about living ethically*. New Haven, CT: Yale University Press.

Spohn, W. (2002). The many facets of the theory of rationality. *Croatian Journal of Philosophy*, *2*(3), 249–264.

Taylor, C. (1975). *Hegel*. Cambridge, England: Cambridge University Press.

Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Cambridge, England: Cambridge University Press.

Tuomela, R. (2013). *Social ontology: Collective intentionality and group agents*. Oxford, England: Oxford University Press.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wallace, R. J. (2018). Practical reason. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/archives/spr2018/entries/practical-reason/

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://lccn.loc.gov/2020048455