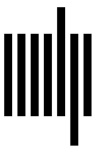
The Handbook of Rationality

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

10.2 Collective Rationality

Hans Bernhard Schmid

Summary

"Collective rationality" is used in a variety of ways in the relevant literature, two of which stand out: (a) the rationality of cooperation and coordination and (b) corporate rationality (or the rationality of group agents). These two conceptions are markedly different from each other in the issues they concern and in the concepts of collectivity and rationality they involve. Upon closer inspection, however, it turns out that from different sides, they converge on the same issue, namely, the capacity for rational joint action.

1. Rational Joint Action, Cooperation, and Corporation

In the first of the two versions in which collective rationality is currently discussed, it is a relational feature of choices of individuals. The paradigmatic case of this sense of collective rationality is mutual cooperation in a two-person prisoner's dilemma. Individual choices are collectively rational if they realize a mutually cooperative outcome (an outcome that is better for both than mutual defection), even though individual strategic reasoning recommends mutual defection. Defection is individually rational (in a strategic sense) because from the point of view of each of the participating individuals, choosing to defect realizes a better state no matter whether the other cooperates or defects. Among strategically rational individuals, mutual defection is thus the expected outcome—which seems a bit foolish given that they could have realized an outcome that is better from both participants' points of view. Collective rationality in this sense is thus defined in contradistinction against (a strategic conception of) individual rationality. Conversely, there is an air of individual irrationality about individual choices that are collectively rational, even though in terms of efficiency, the participants are ultimately better off through mutual cooperation. Collective rationality in this sense is a matter of the rationality of cooperation (although, as will be argued below, *coordination* might be a better paradigm to study this first conception of collective rationality). The rationality of cooperation and coordination matters because it explains the provision of common and public goods (cooperation) and the role of conventions in action (coordination).

In the second salient sense of the word, collective rationality is an attribute of collective or group agents, their attitudes, and their actions. It is the feature in virtue of which collective or group agents (such as political parties, universities, or business corporations) are consistent in their attitudes and actions over more or less complex interdependent decisions. In this second sense, collective rationality is defined in analogy to (rather than in contradistinction against) individual rationality. Just as rational individual agency needs some degree of consistency across issues, it seems that at least some collectives need to ensure that their collective views and decisions are mutually consistent too. The problem of collective rationality in this second sense, as discussed in the recent literature, is that it is not guaranteed by the rationality of individuals. No plausible fixed aggregation procedure robustly ensures the rationality of collective decisions among rational individuals. Achieving collective rationality is thus an extra achievement necessary for organized social life—the "collectivization of reason," as it is sometimes called in the literature. This second conception of collective rationality matters insofar as the credibility and trustworthiness of group agents such as institutions and organizations depend on their consistency in their attitudes and actions.

These two conceptions of collective rationality—the rationality of coordination and cooperation, as well as corporate rationality—thus seem to be markedly different in their targets and approaches. Perhaps more important, they radically differ both with regard to the idea of rationality as well as with regard to the notion of collectivity they entail. The rationality of cooperation between individuals is a matter of *efficiency*, and it involves collectivity in the sense that it is a *relational feature of participant individual choices*. By contrast, the

rationality of collective agents (or corporate rationality) is a matter of *consistency*, and it involves collectivity in terms of *agency that is collective rather than individual*. A further distinction is that while the rationality of cooperation is defined in *contradistinction* from individual rationality, corporate rationality is defined in *analogy* to individual rationality.

After overviews over each of these two debates (sections 2 and 3), this chapter makes a suggestion concerning how, despite their differences, they might hang together (section 4). The claim is that the feature that rationalizes cooperation (and coordination) between individuals is also the basic force that drives toward coherence between group attitudes. Both conceptions thus converge on the capacity to think and act together, from a joint point of view, and to do so consistently and effectively. This capacity is currently mostly discussed under labels such as "collective intentionality."

2. The Rationality of Coordination and Cooperation

In the first sense of the word, it is sometimes claimed in the literature that human social activity cannot be based on individual rationality alone but has to be complemented with—or perhaps partially replaced by—collective rationality. The relation between individual and collective rationality is thought to be difficult because the two can—and often do—come into conflict with each other. The basic problem has been noticed at least since Hobbes (cf. Gauthier, 1979; with a bit of good will, one could perhaps reconstruct a passage in Plato's *Republic* [359a] along similar lines). In a game-theoretic framework (cf. chapter 9.1 by Albert & Kliemt, this handbook), the issue at stake has been at the focus of social theory since the 1950s under the name of "the prisoner's dilemma" (Kuhn, 2017; an alternative—and equally correct—spelling is "prisoners' dilemma").

The prisoner's dilemma (PD) is a situation of interdependent choices between two available options (choices, or strategies), *c* and *d*, where in the case of two participants *A* and *B*, the outcome if *A d*'s and *B c*'s realizes best what *A* values and worst what *B* values (and vice versa for *B*'s *d*'ing and *A*'s *c*'ing), and where mutual *c*'ing realizes what *A* and *B* value better than mutual *d*'ing. The PD does not presuppose selfish preferences or values. It can arise between altruists, too (cf. Tilley, 1991), although it is usually illustrated with the example of self-interested (and mutually disinterested) participants and has its most important applications in this context. The reason why the system of evaluations that is the PD is so important in the debate is that it captures the basic structure of the problem of common and public goods—that is, goods that are nonexcludable and therefore invite free-riding (the "tragedy of the commons"). The dilemma, as it is usually conceived of in the literature, is "a conflict between individual and collective rationality" (Orbell & Wilson, 1979, p. 411). It is claimed that in situations such as the PD, "individually rational" choice results in "collective irrationality" (e.g., Sorensen, 2004, p. 265). In this sense, collective rationality is thus a property of a set of individual choices that are such that the individuals reap the benefits of mutual cooperation in the face of temptation to defect. Because of this structure of incentives, there is, however, an air of irrationality about cooperative choices. Standard strategic rationality demands of individuals that if there is an option that realizes a better outcome (measured by one's own desires or values) no matter what the relevant others do, one should choose that strategy. The cooperative choice is strictly dominated by the defective choice—to choose the cooperative strategy realizes a worse outcome no matter what the other does. This recommends defection in PDs as the rational choice (and since this is true for both participants, mutual defection is the only Nash equilibrium).

It is hard to accept from an intuitive point of view, however, that it should be rational for agents with common knowledge of rationality as well as common knowledge of the payoff structure that they should knowingly realize a combination of choices that will make each of them worse off than another available outcome. To put this point in the relevant jargon: the problem is that mutual defection is the only Pareto-inefficient outcome of a PD. It seems counterintuitive to say that it is rational to be inefficient. Agents whose rationality is only strategic, but foreseeably inefficient, thus appear to be "rational fools" (Sen, 1977).

Although not all relevant authors use the term "collective rationality" in this context, the project of safeguarding (combinations of) cooperative individual choices in PDs from the allegation of irrationality has featured prominently in the agenda of social theory and social philosophy, especially since the 1970s (for an extended discussion see Townley, 2008, chapter 9). The reason why this is an important issue is that there are limits to aligning individual strategic rationality with the production of public goods via institutional design. To some degree, social order depends on individual cooperative-mindedness, and among likeminded people, cooperative-mindedness is efficient—and is thus not simply irrational, although the rationality involved here is somewhat in tension with practical rationality of the standard individual (strategic) kind.

Most prominent among the more recent attempts is Paul Weirich's (2009) conception of collective rationality. Weirich defends participant cooperative choices from the allegation of irrationality by arguing that self-support rather than utility maximization is the standard to which it should be held; a combination of cooperative choices is self-supporting. In this vein, Weirich argues that it is a sufficient condition for the rationality of a combination of choices (a "group act," in Weirich's terms) that its individual members' acts are rational (in his revised sense).

Another conception of collective rationality in this sense is Christopher McMahon's (2001). According to his view, it is (collectively) rational to contribute to a "cooperative scheme" if the "outcome of the scheme, when one's contribution is added to the others that will actually be made, exceeds the value to one of the noncooperative outcome" (McMahon, 2001, pp. 21–22).

This conception of collective rationality is, in turn, similar to David Gauthier's (1986) classical concept of constrained maximization (cf. chapter 12.1 by Fehige & Wessels, this handbook, as well as a considerable number of similar attempts to vindicate the rationality of cooperative choices). However, Gauthier avoids using the term "collective rationality." Indeed, it is far from obvious why the rationality of cooperative choices, as conceived in this debate, should be *collective* rather than individual. After all, it is the *individuals* who profit from mutual cooperation (or suffer from mutual defection), not some collective entity such as "the team." In this regard, one might even think that rather than a *dilemma* (i.e., a choice between equally problematic alternatives), the PD is really a *paradox* (i.e., a self-contradiction; an example for this interpretation is Davis, 1977) or perhaps a *tragedy* (a contradiction between intention and effect; cf. Tilley, 1994). Strategic rationality (that is aimed at maximizing expected utility) recommends a choice of which it is known that it fails to maximize utility among strategically rational individuals. The paradox of the PD is thus, in this view, that "so-called irrational players . . . will both fare much better than so-called rational ones" (Luce & Raiffa, 1957, p. 96; the fact that conversely, "rational" players fare worse than irrational ones might be called the tragedy of the PD).

Although the efficiency of mutual cooperation is certainly a good ground to claim *some sort* of rationality for the respective choices, considered in combination, the sense in which the rationality of cooperation is collective rather than individual certainly needs clarification. From this perspective, rather than pitting collective rationality against individual rationality, the PD seems to separate efficiency from strategy dominance, and it is not obvious how exactly the rationality of the efficient choice involves collectivity other than in the rather obvious sense that it takes at least two to realize a mutually cooperative result.

In this regard, the debate on *coordination* is illuminating. While cooperation in PDs requires the participant to choose against their strategic self-interest, there is no such conflict in pure coordination problems (for a discussion of coordination and cooperation in game theory and behavioral economics see Butler, 2012). In situations of pure coordination, there is no dominant strategy for the participants; what is rational for them to do, individually, depends on what they expect the others to choose (of whom it is known that they are in the same situation). Where none of the available equilibria is better for all of the participants, the problem is usually very effectively regulated by conventions (or salience) in real life, and although it seems intuitively obvious that it is rational for the participants to solve coordination problems by choosing the conventional option, it is impossible to sustain this intuition in terms of individual strategic reasoning (where mutual expectations are interdependent rather than fixed or convergent). Classical rational choice theory cannot recommend the choice of the "salient" strategy as the one rational choice. Even more counterintuitive is the recommendation of classical rational choice theory in the case where one coordination equilibrium is better for all than the other(s) (such as in the Hi–Lo game, where the choice of "Hi" by both participants realizes a better result for both than the choice of "Lo" by both partners; "Lo"/"Lo" is still better for both than "Hi"/"Lo" or "Lo"/"Hi"). In this case, it seems intuitively very clear that common knowledge of rationality and of the utilities rationalizes the according choice (i.e., "Hi"). Choosing "Lo" would be irrational. However, this intuition is, again, not sustained by classical rational choice theory. (Especially counterintuitive is the mixed-strategy equilibrium strategy with a higher probability of the strategy that has the worse coordination equilibrium as a potential outcome, i.e., "Lo." The mixed-strategy equilibrium thus equalizes the expected utilities of the two strategies.)

The general structure of interdependent expectations in situations of coordination ("multiple contingency") leads such social theorists as Talcott Parsons and Niklas Luhmann to think that social science should be framed in systems-theoretic terms rather than in terms of rational agency—a strike through the Gordian knot of rationalizing coordination that seems to have lost a great deal of its appeal in social science over the past couple of decades

(Schmid, 2011; for the use of game theory in the social sciences, cf. chapter 10.4 by Raub, this handbook). In recent action and decision theory, the issue has been taken up, and a series of theories of the rationality of coordinated choices has been developed. In this context, the *collectivity* of the rationality in question has received particular attention. Most prominently, Robert Sugden (1993) and Michael Bacharach (2006) have suggested that the rationality of coordinated choices is collective in the sense of a certain kind, or mode, of reasoning. Rather than reasoning strategically from their individual points of view, the participants engage in *team reasoning*. They recognize situations of coordination (such as the paradigmatic Hi–Lo game) as situations in which "the team"—"we," from the agents' perspective—has to choose, and this rationalizes their contribution to the best collective choice. The way this matters to the prisoner's dilemma is that aversion against exploiting others (or being exploited by them) transforms a PD into a Hi–Lo game with mutual cooperation as the better equilibrium (Gold & Sugden, 2007, pp. 288ff.). The way in which this relates to the open question of the collectivity of the collective rationality in rational team reasoning is that it is "as a team" (rather than "as individuals") that the participants reason. As Susan Hurley (2003) puts it in her pithy title, "The limits of individualism are not the limits of rationality." It is widely accepted in the relevant literature that this should not be taken as saying that it is simply another rational choice for individuals to identify with teams and reason from a team perspective, as this would lead back into the problem of interdependent expectations (it is rational *if* the others choose to identify with the team too, etc.). At this point, Elizabeth Anderson's "priority of identity" principle is relevant: before the question of rational choice can be answered, the question of the identity of the agent has to be settled, and very few agents are "merely individuals" but rather members of a wide variety of groups (Anderson, 2001, p. 30). However, the claim that the—singular or plural—identity of the chooser is always a matter of a subpersonal framing process that is beyond the scope of reasoning is equally dissatisfying as the "rational identity choice" view. Clearly, we do sometimes reason about from which point of view to approach a decision, such as in the case of role conflicts, and it is not clear that we should take a "purely individual" point of view as the default.

3. The Rationality of Corporations

In a second and apparently quite different sense of the word, "collective rationality" is used in the literature for a particular relation between beliefs, desires, and perhaps intentions had by a collection, or group. Collective rationality is the feature in virtue of which there are "consistent and complete group attitudes towards the propositions on the agenda" (List & Pettit, 2011, p. 49; cf. chapter 10.1 by Dietrich & Spiekermann, this handbook). In this sense, the very idea of collective rationality is predicated on the assumption that besides individuals, collectives, too, can have an agenda, that is, that they, too, can be epistemic and practical agents (and be usefully treated as such in social science) and thus be rational. This assumption has received increasing support in the recent literature, most notably from Christian List and Philip Pettit (2011). This marks a decided departure from the widespread older methodological precept in social science and social philosophy to treat only individuals (and not collectives) as agents and thus as (potentially) rational.

On this line, collective rationality is approached as an *aggregation problem*: how can individual judgments (preferences, beliefs, or desires) be aggregated in such a way as to yield a *consistent* collective standpoint across interdependent issues? It turns out on this line of research that no simple nondictatorial aggregation rule ensures collective rationality in a robust way (i.e., across a larger number of interdependent decisions)—and that if we manage to be robustly rational (consistent in our joint position) as the teams and organizations we are, we cannot achieve this through mechanical application of simple judgment aggregation procedures (e.g., majority voting) alone.

This second concept of collective rationality is best introduced indirectly. For in the 20th-century literature in economics and other social sciences, collective rationality—in the sense of rational group agency—has repeatedly and prominently been declared to be neither possible nor even desirable. A particularly clear statement to this effect is in the chapter on individual and collective rationality in James M. Buchanan's (1962/1999) *Calculus of Consent*:

> It is difficult to understand why group decisions should be directed toward the achievement of any specific end or goal. Under the individualistic postulates, group decisions represent outcomes of certain agreed-upon rules for choice after the separate individual choices are fed into the process. There seems to be no reason why we should expect these final outcomes to exhibit any sense of order which might, under certain definitions of rationality, be said to reflect rational social action. Nor is there reason to suggest that rationality, even if it could be achieved through appropriate modification of the rules, would be "desirable." Rational social action, in this sense, would seem to be neither a positive prediction of the results that might emerge from group

> activity nor a normative criterion against which decision-making rules may be "socially" ordered. (pp. 31–32)

In this view, the idea of collective rationality appears simply as an "illegitimate transfer from the individual to society" (Arrow, 1963, p. 120; cf. Suzumura, 1983, p. 94). According to Buchanan's "individualistic postulate," only individuals have a will and values and make choices. Why deny collectives such features and abilities? The worry behind the individualistic position seems to be that collective agency demotes the participating individuals from their place as proper agents to mere "organs" of the collective organism (cf. Buchanan, 1962/1999, pp. 11ff.). This worry is intimately connected to the history of methodological individualism (Heath, 2015; Udehn, 2001). Buchanan's views on collective rationality thus reflect Max Weber's (1922/1968) canonical statements and echo Joseph Alois Schumpeter's (1908/2010, chapter 6) view—with the remarkable difference that both of these classical methodological individualists saw the "individualistic principle" (Schumpeter) as a *methodological* precept for the particular type or branch of social science they envisaged, and they explicitly allowed for conceptions of "societal values" (Schumpeter) or "collective agents" (Weber) for other social sciences.

The most influential challenge to Buchanan's ban on collective rationality comes from Kenneth J. Arrow: "Collective rationality in the social choice mechanism is not . . . an illegitimate transfer from the individual to society, but an important attribute of a genuinely democratic system capable of full adaptation to varying environments" (Arrow, 1963, p. 120). In particular, collective rationality is the remedy against "democratic paralysis"—"a failure to act due not to a desire for inaction but an inability to agree on the proper action" (Arrow, 1963, p. 120). Thus, contrary to Buchanan's view, collective rationality may well appear as desirable, and very desirable indeed. But is it possible?

The way the issue has historically been approached is in terms of *aggregation* of individual preferences to a social preference, and the debate revolved around Condorcet's famous paradox (see Gehrlein, 2006) as the central challenge to collective rationality from an aggregative point of view. Assume a group consisting of *A*, *B*, and *C*. *A*'s preference order is $x > y > z$, *B*'s is $y > z > x$, and *C*'s is $z > x > y$. Each individual member's preferences are complete and transitive. However, if they vote pairwise over the alternatives, it comes out that there is a majority of two for $x > y$, a majority of two for $y > z$, but also a majority of two for $z > x$. Thus, the group's preferences are intransitive: the preference order is circular. Their predicament, as a group, is thus such that based on their aggregation procedure according to the majority rule, they cannot agree on one option, but neither are they indifferent between the options. A further reason to think that this makes the group irrational is that they can now be money-pumped (cf. chapter 8.1 by Grüne-Yanoff, this handbook): give them *x* for free, then offer them *z*, which they prefer over *x* and are thus willing to pay for, then make them pay for *y*, then *x*, then *z* again, and so on and so forth until they are broke. It seems intuitively irrational to have a structure of preferences that makes one vulnerable to such exploitation.

Kenneth J. Arrow's (1963) famous *impossibility theorem*—which can be seen as a generalization of the Condorcet theorem (List, 2011)—shows that collective rationality is incompatible with other plausible conditions of collective choice: unrestricted domain (there is a unique and complete ranking of collective choices no matter what the individual preferences are that go into the decision process), non-dictatorship (no individual has the power to always determine the group's choice), independence of irrelevant alternatives (the social preference between *x* and *y* should depend only on the individuals' preferences over *x* and *y*), and Pareto efficiency (if every participant individual prefers *x* over *y*, there is a collective preference of *x* over *y*). Many authors have searched for—and suggested—potential escape routes. Most prominent among these is Amartya Sen's (1969) proposal of a weaker conception of collective rationality; however, this condition turned out to contain an oligarchy (see Blair, Bordes, Kelly, & Suzumura, 1976).

While Arrow's impossibility theorem concerns preference aggregation, *judgment aggregation* is at the center of the more recent debate on the desirability and (im)possibility of collective rationality. This development might be seen as a generalization (cf. Nehring, 2003), as a shift of focus (depending on whether or not one interprets preferences as judgments—a preference of *x* over *y* can be interpreted as the judgment that *x* is better than [or preferable to] *y*), or perhaps as not making much of a difference at all (if one interprets epistemic judgments as preferences over what to believe). Again, the point is that groups and other collective bodies often have to arrive at judgments concerning what is the case and what should be done, and they have to show some degree of consistency in their judgment in order to function as the group or collective body they are.

For this new wave in the discussion on social choice, the *discursive dilemma* plays a similar role as did Condorcet's paradox for the Arrovian line and has largely replaced the latter in the recent debate. The main point,

however, is the same: the combination of majority voting rule and rationality of the participant individuals does not ensure collective rationality. Philip Pettit's (2003) example for the discursive dilemma is a political party of three members that is unanimously committed to the view that deficit spending is not a good idea. Now a series of policy decisions comes up. The party first votes on whether or not taxes should be increased—two of the three members vote "no," so this is the party view. Then the issue of defense spending comes up—two of three members vote for an increase, one for a reduction, making "increase" the party line. Then the issue of other government spending comes up—again, increase has a two-thirds pro majority, with only one member voting for reduction. Given the party's view on deficit spending, the party is thus inconsistent. However, the participant individuals may well be individually consistent and thus rational: one of the votes for an increase of defense spending and other spending comes from the member who voted for a tax increase, while the second vote for increased defense spending comes from a member who voted for a reduction of other spending, and the second vote for an increase of other spending comes from a member who judged that defense spending should be reduced, so that each of the participants individually respects the incompatibility of increased spending, nonincrease of taxes, and rejection of deficit spending.

Thus, the discursive dilemma again shows that the combination of individual rationality and majority voting does not ensure collective rationality. It is controversially discussed in the literature on the discursive dilemma, however, how much collective rationality really matters. For example, in the context of the question of democratic legitimacy, Fabienne Peter (2009) distinguishes reason-responsiveness from rational consistency (cf. chapter 2.1 by Broome, this handbook) and argues that the latter is not in itself necessary for legitimacy:

> Is it really the case that, as a democratic collective, we ought to satisfy some requirements of collective rationality? Does the binding force of legitimate decision hinge on whether or not these decisions satisfy requirements of rationality? . . . In light of the distinction between reasons and rationality, it is possible to see that to deny that democratic legitimacy includes a rationality requirement does not entail that the demand to respect a democratic decision as legitimate does not depend on reasons. It is only to deny that there is independent normative force in the fact that collective decisions satisfy some conditions of consistency. (pp. 144–145)

While this seems plausible as far as the question of democratic legitimacy is concerned, an issue of credibility and trustworthiness comes up when inconsistency is not plausibly and robustly precluded. Philip Pettit thus argues for the desirability (and indeed necessity) of collective rationality with his example of a political party: "the party cannot tolerate collective inconsistency, because that would make it a laughing-stock among its followers and in the electorate at large; it could no longer claim to be seriously committed to its alleged purpose" (Pettit, 2003, p. 178). Examples for the importance of (this sense of) collective rationality include cases where groups are trusted (be it by their members or by outsiders, e.g., as testifiers); in such cases, it does indeed seem plausible to assume that groups need to achieve, at the collective level, the sort of robust coherence that characterizes rational individual agency. For a group testimony that p to function as such, there has to be trust in collective rationality. One might object, however, that any such argument presupposes what is claimed to be shown, that is, that group agency matters (for an instructive discussion, cf. Roth, 2014).

In an Arrovian vein, List and Pettit (2011) show that collective rationality is incompatible with the counterparts of Arrow's conditions for preference aggregation (cf. List, 2011; cf. chapter 10.1 by Dietrich & Spiekermann, this handbook). The upshot is—very roughly—this: if the participants can have a rational view on everything that is at stake, and if all the individuals' views are to count for the collective judgment, there is no simple aggregation rule that robustly ensures collective rationality. This severely limits the prospects for the sort of coherence that seems necessary to take a group seriously as a rational agent. (Similar to Arrow's theorem, one might object that the simple majority rule works rather well in most cases of collective decision making and that the problem with the simple majority rule arises only in the special case of interdependent decisions; however, a plausible counterargument is that trusting a group requires some modal robustness and that a group that cannot ensure collective rationality in the complicated case cannot be trusted in simple cases.)

The debate on social choice has thus again come to disagree with Buchanan's "individualistic principle": it is argued that collective rationality is desirable after all. However, the result so far is only grist to Buchanan's mill in that it appears that under the stated conditions, no application of a simple aggregation rule will robustly ensure collective rationality, even if all the individual inputs fed into the process are rational.

One way to interpret this result is that collective rationality cannot be conceived of in terms of an aggregation function. Collective rationality is not a matter of organizational design. Collective rationality cannot be achieved reliably by feeding rational individual inputs

into a mechanism that yields a collective attitude. Admittedly, some groups do need such organizational mechanisms—think of modern democracy. (And it is no coincidence that modern social choice theory originates from investigations of voting procedures in mass society.) But what the impossibility results have shown might just be that thinking of collective rationality on this paradigm is problematic or even wrong-headed. Perhaps such organizational measures as voting procedures are attempts to emulate collective rationality under unfavorable circumstances. Smaller groups typically resort to votes only if in the process of deliberation dissent proves to be persistent; in many other cases, groups work out a joint view in the same way they might jointly prepare a meal—not by mechanically assembling whatever the participants happen to bring along but by working out a joint plan and a division of tasks. Thinking about collective rationality on the model of voting mechanisms might be mistaking the emulation for the original.

Indeed, in the last chapter of their book on *Group Agency* (2011, chapter 9), List and Pettit sketch a "behavioral" alternative to the "organizational" route out of the aggregation problem. Roughly speaking, the view is that for groups to get their act together, they have to self-identify with the team in a way that is "immediate" in the sense that it leaves no "identification gap." An identification gap opens up whenever members think of their group in third-personal terms rather than adopting the first-person-plural perspective (e.g., by my thinking of my professional group in terms of "the philosophy department" rather than in terms of *us*). If, from my perspective, the consistency problem is "the philosophy department's," it is not *my* problem. Thus conceived, the group is another agent, over and above the heads of its members, and collective irrationality seems to be a problem only if I am, for some additional reason, bothered by it. From a first-person-plural perspective, however, inconsistency is *our* problem, and it involves me in a way that does not run through any extra reason for me to be bothered by it.

4. Collective Intentionality

Even though these two conceptions of collective rationality differ in the sort of practical problems they concern as well as in the notions of rationality and collectivity involved in each of the conceptions, they point toward one and the same issue: the two solutions to the two different problems of collective rationality converge on the topic of thinking and acting as a team. The sort of reasoning behind collective rationality appears to be collective in a way in which the participants are self-identified as members that is not in need of further individual motivational or evidential support. But how could individuals ever be "blindly" identified with their teams in such a way that their group's interests and consistency requirements should motivate (and perhaps even justify) their contribution without an extra individual motivation?

One way to address this issue is via the first-person-plural perspective. In recent philosophical research, this has mostly been investigated under the label *collective intentionality*, that is, "the power of minds to be jointly directed at objects, matters of fact, states of affairs, goals, or values" (Schweikard & Schmid, 2013). Although philosophers disagree over how to analyze collective intentional states, the first-person-plural perspective figures prominently in this debate. From the members' points of view, their team's attitude (or lack thereof, or inconsistency thereof) is not just some other agent's business, which concerns them only as external constraints on their own actions. Rather, it is *theirs*, collectively. Put in first-personal terms, that is why it concerns *me*, as a member, in a way that is different from a third-party interest or consistency problem, and much closer to the way in which *my own* interests and consistency problems concern me. The way I am committed by my interests, and bothered by my inconsistencies, is via the particular way I know them, which is self-knowledge, self-consciousness, or self-awareness. One way to cash out the capacity for joint thought and action further is thus in terms of plural (prereflective) self-awareness (Schmid, 2014, 2018). The way in which team reasoners and self-identified members know, or feel, to be reasoning together self-commits to a shared goal in the same way in which an individual preference commits one to an individual goal, although the cases are not identical. In the individual case, an attitude of the form "*x* realizes best what I value, but why should I do it?" makes little rational sense (although it may occur in the case of weakness of will): judging that an alternative is best is to be committed to choosing it in the same way as judging that something is true is to believe it; there is, under normal circumstances, no room for questions concerning whether or not a proposition should be believed after its truth has been settled. In the collective case, however, judgment and rational choice seem to come apart, to some degree. To judge, as a team, that *x* is best *for us* leaves some room open for rational questioning whether or not *I* should contribute (one might need some assurance from others before deciding to pull one's weight). It should be noted, however, that this plural case is still very different from a third-personal case.

"Our" judgment commits me to a different degree than my own, but in contrast to an outsider's judgment, it *does* commit me. If I decide to act against what we, together, judge to be best, I certainly owe an explanation, while the converse is not true: our judgment provides a *pro tanto* reason for my contribution (cf. Gilbert, 2013). After all, the view in question is *ours*, and that implicates me—it is first-personal, although not in the singular form, but in the plural. Thus, our judgment self-commits me (I need no reason to be committed by a judgment besides knowing it to be ours) in the same general way as my own judgment self-commits me just in virtue of my self-awareness of it.

This also bears on the question of consistency, and again, the analogies and disanalogies to the singular case are instructive. Our individual minds contain inconsistent attitudes aplenty, but becoming *aware* of them (in the right first-personal way) is being committed to sorting them out. I can perhaps *have* incompossible intentions, but as soon as I become first-personally *aware* of them, they lose their status as practical commitments, and the issue concerning what to do is reopened. Plural self-awareness works in the same way: it is in virtue of our first-person-plural knowledge of attitudes as *ours* that we cannot be unbothered by their potential inconsistencies but have to make up our (plural) mind about what to think and what to do. In this way, plural (prereflective) self-awareness constitutes a *collective* consistency requirement in the same way as singular (prereflective) self-awareness constitutes an *individual* consistency requirement. In this way, collective self-awareness makes it not only desirable but possible, too, to be effective and consistent as the teams we are.

References

Anderson, E. (2001). Unstrapping the straitjacket of "preference": A comment on Amartya Sen's contributions to philosophy and economics. *Economics & Philosophy*, *17*, 21–38.

Arrow, K. J. (1963). *Social choice and individual values*. New York, NY: Wiley.

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory*. Princeton, NJ: Princeton University Press.

Blair, D. H., Bordes, G., Kelly, J. S., & Suzumura, K. (1976), Impossibility theorems without collective rationality. *Journal of Economic Theory*, *13*, 361–379.

Buchanan, J. M. (1999). *The calculus of consent*. Indianapolis, IN: Liberty Fund. (Original work published 1962)

Butler, D. J. (2012). A choice for "me" or for "us"? Using we-reasoning to predict cooperation and coordination in games. *Theory and Decision*, *73*, 53–76.

Davis, L. H. (1977). Prisoners, paradox, and rationality. *American Philosophical Quarterly*, *14*(4), 319–327.

Gauthier, D. (1979). *The logic of Leviathan: The moral and political theory of Thomas Hobbes*. Oxford, England: Clarendon Press.

Gauthier, D. (1986). *Morals by agreement*. Oxford, England: Clarendon Press.

Gehrlein, W. V. (2006). *Condorcet's paradox*. Berlin, Germany: Springer.

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

Gold, N., & Sugden, R. (2007). Theories of team agency. In F. Peter & H. B. Schmid (Eds.), *Rationality and commitment* (pp. 280–312). Oxford, England: Oxford University Press.

Heath, J. (2015). Methodological individualism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/archives/spr2015/entries/methodological-individualism/

Hurley, S. (2003). The limits of individualism are not the limits of rationality. *Behavioral and Brain Sciences*, *26*(2), 164–165.

Kuhn, S. (2017). Prisoner's Dilemma. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/archives/spr2017/entries/prisoner-dilemma/

List, C. (2011). The logical space of democracy. *Philosophy & Public Affairs*, *39*(3), 262–297.

List, C., & Pettit, P. (2011). *Group agency*. Oxford, England: Oxford University Press.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. New York, NY: Wiley.

McMahon, C. (2001). *Collective rationality and collective reasoning*. Cambridge, England: Cambridge University Press.

Nehring, K. (2003). Arrow's theorem as a corollary. *Economic Letters*, *80*, 379–382.

Orbell, J. M., & Wilson, L. A., II (1979). Institutional solutions to the *N*-prisoners' dilemma. *American Political Science Review*, *72*(2), 411–421.

Peter, F. (2009). Political legitimacy without collective rationality. In B. de Bruin & C. F. Zurn (Eds.), *New waves in political philosophy* (pp. 143–157). New York, NY: Palgrave.

Pettit, P. (2003). Groups with minds of their own. In F. Schmitt (Ed.), *Socializing metaphysics* (pp. 167–195). New York, NY: Rowman & Littlefield.

Roth, A. S. (2014). Indispensability, the discursive dilemma, and groups with minds of their own. In S. R. Chant, F. Hindriks, & G. Preyer (Eds.), *From individual to collective intentionality: New essays* (pp. 137–162). Oxford, England: Oxford University Press.

Schmid, H. B. (2011). The idiocy of strategic reasoning. *Analyse & Kritik*, *1*, 35–56.

Schmid, H. B. (2014). Plural self-awareness. *Phenomenology and the Cognitive Sciences*, *13*, 7–24.

Schmid, H. B. (2018). The subject of "we intend." *Phenomenology and the Cognitive Sciences*, *17*, 231–243.

Schumpeter, J. A. (2010). *The nature and essence of economic theory*. New Brunswick, NJ: Transaction Publishers. (Original work published 1908)

Schweikard, D. P., & Schmid, H. B. (2013). Collective intentionality. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/archives/sum2013/entries/collective-intentionality/

Sen, A. K. (1969). Quasi-transitivity, rational choice and collective decisions. *Review of Economic Studies*, *36*(3), 381–393.

Sen, A. K. (1977). Rational fools. *Philosophy & Public Affairs*, *6*, 317–344.

Sorensen, R. (2004). Paradoxes of rationality. In A. R. Mele & P. Rawling (Eds.), *The Oxford handbook of rationality* (pp. 257–277). Oxford, England: Oxford University Press.

Sugden, R. (1933). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy & Policy*, *10*, 69–89.

Suzumura, K. (1983). *Rational choice, collective decisions, and social welfare*. Cambridge, England: Cambridge University Press.

Tilley, J. (1991). Altruism and the prisoner's dilemma. *Australasian Journal of Philosophy*, *69*(3), 246–287.

Tilley, J. (1994). Accounting for the "tragedy" in the Prisoner's Dilemma. *Synthese*, *99*(2), 251–276.

Townley, B. (2008). *Reason's neglect: Rationality and organizing*. Oxford, England: Oxford University Press.

Udehn, L. (2001). *Methodological individualism: Background, history and meaning*. London, England: Routledge.

Weber, M. (1968). *Economy and society*. Berkeley: University of California Press. (Original work published 1922)

Weirich, P. (2009). *Collective rationality: Equilibrium in cooperative games*. Oxford, England: Oxford University Press.

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