

Collective, universal, and joint rationality

Paul Weirich

Received: 6 December 2005 / Accepted: 4 May 2007 / Published online: 5 July 2007
© Springer-Verlag 2007

Abstract Rationality's extension from individuals to groups yields collective rationality. Just as individuals should be rational, groups should be collectively rational. This paper briefly presents collective rationality and then compares it to universal rationality and joint rationality. Universal rationality is the rationality of all members of a group. It directs individualistic evaluation of a collective act. Joint rationality is the rationality of each individual's part in a collective act given the collective act's realization. Game theory uses it to characterize solutions to games. Collective rationality is not the same as either universal or joint rationality. However, in certain ideal conditions collective rationality agrees with universal rationality, and in other ideal conditions it agrees with joint rationality. Distinguishing the three types of evaluation and explaining their relations contributes to a general theory of rationality.

1 Collective rationality

Groups perform acts. A committee passes a resolution, for instance. Is the committee's act rational? A group acts freely in virtue of the freedom of its members. Although a committee lacks a mind, its act is evaluable for rationality because the committee's act is composed of the free acts of its members and because nothing outside the committee

P. Weirich (✉)
Department of Philosophy, University of Missouri, Columbia, MO 65211, USA
e-mail: weirichp@missouri.edu

controls its act.¹ To evaluate the committee's act, one extends principles of rationality to groups. An account of collective rationality guides their extension.²

Collective rationality, as I introduce it, is an extension to groups of the ordinary concept of rationality.³ Because it is a combination of familiar concepts, collective rationality is meaningful even if its principles are controversial. This paper does not analytically define collective rationality. Its sections taken together introduce collective rationality through a variety of examples and general principles. This section treats rationality, the concept extended to reach collective rationality. Rationality is a familiar concept. Rescher (1988) and other authors explain it. I just highlight features of rationality important in its extension to groups.

Rationality's evaluation of an act is sensitive to an agent's circumstances and abilities. Its demands adjust to them. When circumstances are adverse, it lowers its requirements. For example, it may excuse overlooking top options when the time for a decision is short. In common terminology, rationality is bounded. In particular, it is attainable. In every decision problem, some option is rational.

A decision problem requires an agent to adopt an option from a set of options. A rational option is at least as well supported by reasons as any other option is. Typically, not all options are equally well supported, and some options are not rational. Because rationality is attainable, at least one option is rational. It satisfies all applicable principles of rationality. In a simple ideal decision problem, the principle to maximize utility applies. Some option maximizes utility and therefore is rational. In a nonideal problem, the principle of satisficing may apply. If so, some option is satisfactory and therefore is rational even if it fails to maximize utility. For example, a chess player, having imperfect strategic reasoning power, may rationally make a satisfactory but non-optimal move. Although the principles of rationality for nonideal cases are controversial, rationality's attainability follows from the noncontroversial principle that "ought" implies "can".

Because rationality is attainable, some common principles of rationality are best interpreted as presenting goals of rationality rather than requirements of rationality. Realizing an option at the top of one's preference ranking of options is not a requirement of rationality when circumstances are adverse, but it is a goal of rationality. A rational person pursues goals of rationality in a reasonable way even if obstacles make their attainment difficult. Rationality offers the best way of pursuing the primary

¹ Pettit (2001, p. 5) contends that freedom is fitness for responsibility. Although acts evaluable for rationality are free, freedom does not entail responsibility. A young child is free but not responsible. Also, an adult may escape responsibility for a free act performed out of ignorance.

² I treat principles that assess an act for rationality. A bystander may apply the principles after the act has been performed. The agent, to perform a rational act, need not apply the principles to reach a decision to perform the act. Extending the principles to groups affords insights into the nature of rationality. For example, it reveals that an act's being intentional is not necessary for the act's being evaluable for rationality. A group's act is not intentional, but is evaluable for rationality if free and fully controlled by the group's members.

³ Sen (1977) notes the deficiencies of technical definitions of rationality and advocates a circumspect view of rationality. I work within the program he advocates. I forgo technical definitions of rationality and use only the ordinary concept of rationality so as not to miss its nuances and richness.

goal of successful action, subject to constraints the agent's abilities and circumstances impose.

Although rationality does not require utility maximization in all cases, in some cases it does. In those cases an option is rational only if it is utility maximizing given its realization. An act's being rational depends on its having a conditional property. If you do not buy a lottery ticket, then an evaluation of your buying a ticket depends on the act's (expected) utility in hypothetical circumstances in which you buy a ticket. In that respect an act's being rational is like sugar's being soluble. Just as sugar's being soluble requires that it dissolve if placed in water, an act's being rational requires that it maximize utility if performed. Classifying options as rational involves hypothesizing in this way. Although an agent realizes only one option in a decision problem, rationality's principles say for every option whether the option is rational. Rationality classifies all options, not just the one realized, because it is a source of advice about options to realize.⁴

Rationality evaluates some but not all acts by comparisons with alternatives. Take, for example, a sequence of acts such as having dinner and going to a movie. Rationality may evaluate the sequence by evaluating its steps. The sequence may be rational because having dinner is rational and because going to a movie is rational. Rationality need not evaluate the sequence by comparing it with alternative sequences such as skipping dinner and going for a walk. Rationality uses comparisons to evaluate acts over which an agent has direct and temporally immediate control. It uses evaluations of components to evaluate composite acts over which an agent has temporally extended or indirect control.

Standards of evaluation differ in scope. Some standards evaluate an act taking for granted an agent's beliefs and desires. Other standards have wide scope and downgrade an act if it rests on defective beliefs and desires. I call principles with wide scope principles of *comprehensive* rationality. They take none of an agent's features for granted when evaluating an act. They evaluate an act together with its grounding in an agent's beliefs, desires, and deliberations.

Because my project is extending the ordinary concept of rationality to groups, I put aside technical definitions of collective rationality. According to one technical definition, collective rationality is efficiency, or Pareto optimality. In a symmetrical bargaining problem agents may achieve Pareto optimality without achieving an equal division and so without achieving collective rationality. Pareto optimality is not sufficient for collective rationality in such problems. Also, in nonideal circumstances, Pareto optimality is too demanding to be a necessary condition of collective rationality.

Although Pareto optimality is not a requirement of collective rationality, it is a goal of collective rationality. Its being a goal of collective rationality does not mean that a group, if rational, has Pareto optimality as a goal. A goal of collective rationality expresses an achievement of rational ideal agents in ideal circumstances. A feature of

⁴ To evaluate an option for an agent, rationality takes account of the option's causal influence on states of the world, including the acts of other agents. It does not hold fixed probabilities of states of the world if realizing the option would causally influence those states. This is a lesson of causal decision theory, which Joyce (1999) explicates.

a collective act, such as Pareto optimality, is a goal of collective rationality if in ideal conditions for joint action, rational ideal individuals generate a collective act with that feature. This may happen as if an invisible hand directs the group of individuals. Not every individual in the group need form a desire that the group meet the goal. However, fully rational and fully informed ideal agents each desire the goal's realization. The goal motivates them, and they achieve it deliberately in ideal conditions for joint action.

According to another technical definition, collective rationality is social optimality, that is, maximization of social welfare. However, people may have a good excuse for failing to achieve social optimality. For instance, they may not know where social optimality lies and so head in the wrong direction. Collective rationality does not require social optimality in adverse circumstances.

Sen (2002, p. 290) suggests a purely procedural conception of collective rationality. It holds that collective rationality is whatever emerges from sensible social institutions. This view is too tolerant. Sensible social institutions in the hands of irrational individuals may fail to generate collectively rational acts. Collective rationality imposes substantive standards in addition to procedural standards. For example, in ideal cases, such as perfectly competitive exchange economies, collective rationality demands Pareto optimality. Attaining that goal is a substantive requirement in favorable circumstances for joint action.

Theorists use two methods of extending rationality to groups. One method proceeds by analogy. Take, for example, the principle to realize an option at the top of one's preference ranking of options. To apply the principle to a group's act, one may define the group's preference ranking of options and then check whether its act is at the top. The second method of extension evaluates a group's act by evaluating the individual acts that constitute that act. This method uses evaluation by components instead of evaluation by comparisons.

An account of collective rationality that relies on analogy constructs technical definitions of the collective analogues of preference and other mental states with a role in principles of individual rationality. Intuition is not a reliable guide to suitable definitions, as Arrow's Theorem (1951) shows. One must verify that the technical definitions, combined with the principles of rationality, yield acts that are collectively rational. One needs first principles of collective rationality to verify the principles analogy yields.

The most basic type of evaluation of a collective act examines the act's components. It declares a group's act to be rational if it is composed of rational acts by the group's members. For a collective act, evaluation by components is more fundamental than evaluation by comparisons because evaluation by components does not rely on technically defined collective analogues of an individual's mental states.

Principles of individual rationality that evaluate an act by comparisons with rival acts may apply to groups given suitable collective counterparts of mental states. Such evaluations may agree with evaluations by components. However, principles using evaluation by comparisons operate reliably only in a restricted range of cases. For example, the principle to pick an option that maximizes expected collective utility works reliably only if a group's members make the same probability assignments, as Broome (1987) observes.

Action theory confirms that evaluation by components is the basic method of assessing a collective act. An agent controls some acts directly and other acts indirectly. Rationality guides control over action, and so its fundamental evaluations treat acts that an agent controls directly. A group does not directly control its acts. It performs an act only through the acts of its members. Hence, evaluating its acts by comparisons with rival acts is a derivative form of evaluation. Rationality's first principles for evaluation of a group's act proceed by evaluating its components. Rationality's fundamental principles evaluate exercises of direct control, and only the individual acts constituting a collective act are products of direct control.

To achieve practical consistency, rationality does not tell each of two chess players to win their game. They cannot both win. Rationality tells each to try to win. Both may try to win. A general theory of rationality must issue consistent directives to a group and to its members. To be consistent, it cannot tell a crowd to sit and also tell each member of the crowd to stand. Evaluating collective acts by components guarantees practical consistency. By taking individual rationality to entail collective rationality, a general theory of rationality ensures the consistency of its standards for individuals and for groups. Individuals and any group they constitute may simultaneously comply with the theory's requirements. In some possible world all parties may resolve their current action problems without violating any principle of rationality.

A common objection to collective rationality's evaluation of collective acts by components emerges in studies of the Prisoner's Dilemma. If the individuals in a Prisoner's Dilemma both defect, they each act rationally but the pair acts irrationally, the objection claims, because the pair fails to achieve Pareto optimality.⁵ This objection conflates goals and requirements of collective rationality. Pareto optimality is a goal but not a requirement of collective rationality. In the Prisoner's Dilemma conditions are not ideal for achieving goals of collective rationality. The individuals cannot communicate and form binding contracts, or take other steps to facilitate joint action. In nonideal conditions, groups may meet requirements of collective rationality without attaining all goals of collective rationality. In a Prisoner's Dilemma, if each individual is rational, the pair is rational despite its failure to achieve Pareto optimality. Collective rationality lowers its demands in circumstances that impede joint action.

A general theory of rationality should not tolerate conflict between individual and collective rationality in the Prisoner's Dilemma. One rightly ignores a theory of rationality that demands the impossible of groups and their members in the Dilemma. A general theory of rationality increases its practical value if it resolves conflicting demands and achieves practical consistency. The theory may do this without ignoring goals of collective rationality. Besides telling individuals to be rational in the Dilemma, it may tell them to design social institutions that prevent the Dilemma from arising. It may urge realization of goals of collective rationality without putting collective rationality at odds with individual rationality.

Rationality's extension to groups is a theoretical enterprise, and its success depends on its theoretical fruitfulness. The extension leaves rationality's native land, and so intuitions about collective rationality are not a complete guide to a fruitful theory. In

⁵ For example, Sen (2002, p. 212) states that in the Prisoner's Dilemma individual and collective rationality conflict.

the foreign territory, theoretical objectives should also direct principles of rationality. The point of rationality's extension to collective acts is to evaluate collective acts and to direct collective action. A conflict between principles of collective rationality and principles of individual rationality thwarts this theoretical objective. A general theory of rationality better achieves its objective if it resolves the conflict. Collective rationality's evaluation of collective acts by their components resolves the conflict and enhances the value of the general theory of rationality. Adopting this method of evaluation, the general theory yields practically consistent guidance to a group and to its members.

2 Universal rationality

Universal rationality offers an individualistic standard for a collective act's evaluation. A group's act achieves universal rationality if and only if the contributions of all members are rational. The rationality of all members is plain and simple. However, a standard of rationality classifies all possible acts in a collective action problem, not just the act realized. It evaluates each act profile, that is, each combination of members' acts with exactly one act for each member. The universal rationality of a profile not realized is open to various interpretations. How should one characterize its universal rationality to express the individualist's standard for groups?

I look for a characterization of universal rationality that makes it agree with collective rationality. Given agreement, universal rationality issues a useful derivative standard of evaluation. One may use it to replicate collective rationality's evaluation of a group's act. Collective rationality issues the basic standard of rationality for groups, but in some cases the derivative standard of universal rationality may be easier to apply. It may offer a short cut to a collective act's evaluation.

Suppose that the members of a group do not all act rationally. Then their all acting rationally, their universal rationality, is hypothetical. Various specifications of the hypothetical situation in which all act rationally yield various interpretations of universal rationality. Does any interpretation align universal rationality with collective rationality?

First, consider a profile of acts that assigns to each individual an act that is rational with respect to actual circumstances. To form such a profile, one identifies for each individual an act that is rational and creates a combination of such acts, one for each individual. That combination of rational acts, if it were realized, need not yield collective rationality. For example, consider the two-person coordination game that Fig. 1 depicts.

Imagine that the game involves two people who would like to meet at one of two places. Suppose that convention establishes the place to meet, and Row and Column go there if they realize the profile (T, L) . On a whim, Row irrationally deviates from

Fig. 1 Coordination

	Left	Right
Top	1, 1	0, 0
Bottom	0, 0	1, 1

the convention and performs B . Fortunately, he signals his intention. Although Row is not aware of it, Column anticipates his performing B and rationally responds with R so that (B, R) is realized. In this example, the combination of acts rational with respect to actual circumstances is (T, R) . It has Row's rational act and Column's rational act. This profile, taken as universally rational, does not agree with collective rationality. In this example, to be collectively rational, a profile must be such that if it were realized, all individuals would be rational. If Row were to act rationally and perform T , Column's rational response would be L , assuming that Column would anticipate Row's act. The profile achieving collective rationality in which Row realizes T is (T, L) and not (T, R) .

In general, suppose that some individual adopts an irrational strategy. If that agent were to act rationally, then the act rational for another agent might change. A profile of strategies in which each agent is rational assigns to the irrational agent a hypothetical strategy. If others anticipate his strategy, then the strategies rational for them in actual circumstances may differ from the strategies rational for them if he hypothetically changes his strategy. A profile in which each strategy is rational with respect to actual circumstances need not be a profile in which all strategies are rational with respect to the hypothetical circumstances of their realization together. A combination of strategies, each of which is rational in actual circumstances, need not achieve collective rationality.

As a second interpretation of universal rationality, consider being a profile that would be realized if all were rational. In a case where all individuals act rationally, this interpretation yields the profile of acts they perform. In a case where some individual acts irrationally, the interpretation yields a profile in which each act is rational when combined with the others. In the sample coordination problem, the profile emerging from the rationality of all is (T, L) . That profile is collectively rational, too.

Under the second interpretation, universal rationality still diverges from collective rationality. Suppose that Row knows that Column anticipates his act, whatever it is. Then the pair is also collectively rational if it realizes the profile (B, R) . That profile does not count as universally rational under the second interpretation. It does not result from the rationality of all. Only (T, L) does. So the second interpretation of universal rationality fails to align it with collective rationality. In general, there is just one profile that would be realized if all individuals were rational. Yet in some situations there are many ways for a group to be collectively rational.

As a third interpretation of universal rationality, consider being a profile of acts such that, if it were realized, all acts in it would be rational. In this case a universally rational profile, if it were realized, would yield a collectively rational profile. Moreover, multiple ways of achieving collective rationality yield multiple profiles counting as universally rational. In the sample coordination problem, if Row and Column communicate to arrange a meeting place, the new interpretation of universal rationality counts both (T, L) and (B, R) as universally rational.

Under the new interpretation, individual acts are universally rational if, were they realized, all would be rational. Universal rationality thus yields an act profile in which all the members of a group act rationally. If all the members of a group act rationally, then their combination of acts is collectively rational. So a profile's achieving universal rationality entails its achieving collective rationality. Universal rationality

is nonetheless not the same as collective rationality because the reverse entailment does not hold. A profile's achieving collective rationality does not entail its achieving universal rationality. A group's act may be collectively rational even if not all members act rationally in realizing it. For example, a committee may elect the best candidate for an office even if not all vote for her. The committee may elect rationally even if some members vote irrationally.

Under the third interpretation, universal rationality comes as close as possible to collective rationality. I adopt this interpretation. Although the interpretation does not make universal and collective rationality agree in all cases, it makes universal rationality entail collective rationality. This is a valuable feature of the interpretation. Under the interpretation, universal rationality is a reliable, although not exhaustive, means of identifying collectively rational profiles.

To compare further universal and collective rationality, consider a question arising in a general theory of rationality. Does universal rationality agree with collective rationality under any standard idealizations? Idealizations control for some factors in the complex set of factors explaining an act's rationality. Perhaps some idealizations extended from individuals to groups bring universal rationality into alignment with collective rationality. A collectively rational group, if free of mistakes, realizes a profile with universal rationality. So perhaps some idealizations eliminating mistakes align universal and collective rationality.

To set the stage for comparing universal and collective rationality under idealizations, consider a preliminary question about collective rationality. What makes a collective act rational when the individual acts constituting it are not all rational? The act may achieve collective rationality by duplicating the outcome of a hypothetical collective act constituted by individual acts that are all rational. The standards of collective rationality evaluate an act by its outcome and recognize that in some cases several combinations of individual acts achieve the same outcome. Just as a person may greet a friend by waving either the left-hand or the right-hand, a committee may elect a candidate with various combinations of votes. Its electing a candidate is a rational collective act if that candidate would have been elected had all members voted rationally. In general, a collective act's equivalence to a universally rational combination of individual acts suffices for its collective rationality.

Comprehensive rationality eliminates irrationalities in the grounds of an act. Although universal rationality and collective rationality are distinct, do their comprehensive counterparts agree? Agreement requires mutual entailment. If each individual is comprehensively rational, then together they achieve comprehensive collective rationality. If a group's act is collectively rational in the comprehensive sense, then perhaps each member's contribution is comprehensively rational, too. Maybe comprehensive collective rationality entails universal comprehensive rationality.

The move to comprehensive standards fails to align universal and collective rationality. Standards of comprehensive rationality generally tolerate inconsequential mistakes.⁶ Comprehensive rationality's evaluation of an individual's act usually accepts

⁶ Weirich (2004, Chaps. 6, 7) explicates comprehensive rationality and explains the mistakes it tolerates.

a mistake in its grounds, such as a mistake in the beliefs that prompt the act, if the mistake is inconsequential. For instance, a spectator who irrationally thinks that one of two adjacent stadium seats is better may in a comprehensive sense rationally choose that seat despite the error in belief, because the error is inconsequential. Similarly, a committee may in the comprehensive sense rationally elect a candidate despite some irrational votes if the mistaken votes are inconsequential. If, despite some member's irrationality, a group achieves an act following from universal rationality, that member's irrationality is inconsequential. So a group's act may be comprehensively rational despite the absence of universal comprehensive rationality.

The idealization of full rationality goes beyond comprehensive rationality. It takes an agent to be rational in all matters. For example, it precludes irrationality in any step of a sequence of acts. For a group, full rationality precludes irrationality in any component of its acts. Given its full rationality, the individual acts that constitute its collective acts are rational. Hence, adopting the idealization of full rationality for a group aligns collective rationality with universal rationality.

Universal and collective rationality's agreement under the idealization of full rationality has theoretical interest but little practical value. It does not assist inferences from collective to universal rationality. Before inferring that a collectively rational outcome is also universally rational, one must first verify the assumption of full rationality, and that requires verifying that the outcome is universally rational. Verifying the assumption that licenses the inference makes the inference superfluous. Although a group's full rationality is distinct from its universal rationality, its full rationality entails its universal rationality.

To complete the comparison of collective and universal rationality, consider their attainability. Collective rationality is attainable. Is universal rationality attainable? Universal rationality's attainability has two interpretations. Under one interpretation, a group may attain universal rationality if it is possible for all the members of the group to act rationally. Under this interpretation, universal rationality is attainable. It is possible not only for each to act rationally but also for all to act rationally. In some games it is not possible for all to win, but nonetheless it is possible for all to play rationally. If all were rational, a certain strategy profile would be realized. That profile achieves universal rationality. If all were rational, they would achieve both universal and collective rationality.

Under another interpretation, universal rationality's attainability requires an act profile such that, if it were realized, all acts in it would be rational. The possibility that all agents act rationally does not ensure such an act profile. Consider the profile that would be realized if all were rational. It may not be the case that if this profile were realized, all would be rational. In some games, for each profile, if it were realized, some player would be irrational. For example, consider the game Matching Pennies, which Fig. 2 presents.

Fig. 2 Matching Pennies

	Heads	Tails
Heads	2, 0	0, 2
Tails	0, 2	2, 0

Suppose that only pure strategies are available, and suppose that each player anticipates the profile realized whatever it is. Then, for each profile, some player fails to maximize utility. If rationality requires utility maximization, each profile's realization entails some player's irrationality. Under the second interpretation of attainability, universal rationality is not attainable.⁷

The example raises doubts also about universal rationality's attainability under the first, individualistic interpretation of attainability. In Matching Pennies, what is the profile realized if all act rationally? If all were to act rationally, and that required maximizing utility, then the players would not each anticipate the other. One player would maximize utility with respect to erroneous information about the other. To fill out the case, imagine that Row would be misinformed about Column and that (H, T) would be realized. Despite first impressions, that supposition does not contravene the conditional that if (H, T) were realized, Row would fail to maximize utility. The law of contraposition does not govern counterfactual conditionals. It may be true that (1) if all were to act rationally, (H, T) would be realized and (2) if (H, T) were realized, not all would act rationally. The second conditional does not entail that (3) if all were to act rationally, (H, T) would not be realized. This observation allays doubts about universal rationality's attainability in the first, individualistic sense. Universal rationality's unattainability in the second, profile sense does not preclude its attainability in the first, individualistic sense.⁸

3 Joint rationality

Joint rationality is a topic of game theory. In a typical game some but not all profiles of players' strategies are jointly rational. The strategies in a profile, one for each player, are jointly rational if and only if they are all rational taken together. Joint rationality is similar to universal rationality. However, its interpretation in game theory distinguishes it from universal rationality.

Solutions to a game may be defined either objectively so that they depend on the game's features or subjectively so that they depend on the players' beliefs about those features. In ideal cases players know their game's features, so objective and subjective solutions agree. Solutions subjectively defined arise from the players' rational decisions. They occur when the players rationally use their beliefs and desires to choose strategies. A solution in the subjective sense is a profile of jointly rational strategies. A profile's strategies are jointly rational if and only if each strategy in the profile is rational given the profile. An explication of joint rationality attends to its role in an account of subjective solutions and their agreement with objective solutions in ideal games.

⁷ In this version of Matching Pennies every outcome is *rationalizable* in the sense of Bernheim (1984) and Pearce (1984), but no profile is *rationalized* if agents anticipate the profile realized. That is, no profile has only strategies that maximize utility given the information the agents have if the profile is realized.

⁸ Weirich (1998) treats the logic of counterfactual conditionals in games such as Matching Pennies and pursues in greater depth the issue of universal rationality's attainability.

von Neumann and Morgenstern (1944, p. 146–47) argue that a theory of rationality specifies a solution to a game that all players may foresee. A solution specifies a strategy for each player that is rational in the situation the profile yields. It is a profile of strategies such that each strategy is rational given the profile.⁹

Nash (1950) elaborates the account of equilibrium a game's solution achieves. A Nash equilibrium is a profile in which each strategy is payoff maximizing given the profile. A subjective version of Nash equilibrium uses utility maximization in place of payoff maximization. Accordingly, an equilibrium is a profile such that no agent prefers to deviate unilaterally. Each agent's participation in the profile is rational given knowledge of the profile. Objective and subjective Nash equilibria agree in ideal conditions under which players know their game's features.¹⁰

The strategies in a solution are rational in a comprehensive sense and do not rest on irrational beliefs or other mistakes. A solution achieves comprehensive joint rationality. In ideal, mistake-free contexts, rationality amounts to comprehensive rationality, so in those contexts strategies that are jointly rational are also jointly rational in a comprehensive sense.

The strategies in a profile are jointly rational if and only if each is rational given the profile. In this characterization of joint rationality, should the profile's realization be supposed as a fact only or as an addition to the agents' knowledge? A profile of jointly rational strategies, taken as a solution, provides strategies that are rational in light of the profile. Each strategy is rational given information about the other strategies. The profile's supposition affects the rationality of strategies only if it affects information. So it appears that the supposition should take the profile as an extra bit of knowledge. This approach works fine in ideal games in which agents figure out each other, as von Neumann and Morgenstern imagine. However, so that the account of solutions extends to nonideal games in which agents have imperfect insight into others' behavior, I define joint rationality in terms of a profile's realization. This approach offers more versatility than defining it in terms of knowledge of the profile's realization. Factors such as the type of supposition and background assumptions may make supposition of the profile yield knowledge of the profile. Because of those factors, in ideal games supposition of a profile carries knowledge of the profile, as desired for an account of solutions.

Next, consider how joint rationality should imagine a profile's realization. Should supposition of a profile's realization preserve evidential or causal relations? Which is the right approach for an account of solutions? This is a delicate issue. I begin by reviewing the distinction between evidential and causal types of supposition.

⁹ Some theorists assert that joint rationality obtains only if each strategy is rational given the other strategies. However, a strategy's rationality may depend not only on other agents' strategies but also on the strategy itself. So it is better to say that joint rationality requires each strategy to be rational given the whole profile and not just given the profile's remainder.

¹⁰ An equilibrium-in-beliefs, in the sense of Aumann (1987), is similar to, but distinct from, a subjective Nash equilibrium. Equilibrium-in-beliefs uses nonconditional utility maximization, whereas subjective Nash equilibrium uses utility maximization given a profile realized. In a typical game potential equilibria-in-beliefs abound, but given agents' beliefs as they are in the game, at most one equilibrium-in-beliefs exists. In contrast, multiple subjective Nash equilibria may exist. Also, in an equilibrium-in-beliefs an agent's strategy has the same probability according to all other agents, and the agent knows their probability assignments. This need not happen in a subjective Nash equilibrium.

Consider a coin with an unknown bias. Getting Heads on a toss furnishes evidence that the next toss will be Heads but does not cause the next toss to be Heads. Evidential and causal relations differ. Types of supposition differ according to ways of fleshing out a supposition. One common type fleshes out a supposition in ways that preserve evidential relations. Another common type fleshes out a supposition in ways that preserve causal relations. Consider these two conditionals. (1) If Shakespeare did not write *Hamlet*, someone else did. (2) If Shakespeare had not written *Hamlet*, someone else would have. The first conditional is true, and the second is false. The two conditionals make the same supposition about Shakespeare but flesh it out differently. The first attends to the evidence that someone wrote *Hamlet*. The second attends to the causal improbability of someone's duplicating Shakespeare's play.

One supposes a proposition p with a world w as background for the supposition. Usually, the actual world furnishes the background for the supposition. A type of supposition may be represented by a selection function s applied to p given w . The value of $s(p, w)$ is the p -world nearest to w . In other words, $s(p, w)$ minimally revises w to accommodate p . In a world w , a conditional "if a then c " is true just in case c is true in $s(a, w)$, that is, the a -world nearest to w , or the minimal a -revision of w . Let e denote a selection function favoring evidence, and let c denote a selection function favoring causation. To obtain $e(p, w)$, one minimally revises the evidential dynamics of w to accommodate p . To obtain $c(p, w)$, one minimally revises the causal dynamics of w to accommodate p . In a typical case, supposing that p generates minor changes in the evidential or causal dynamics of w and mainly changes in the events occurring within those dynamics.

Let w represent the actual world, and let $\sim s$ represent the proposition that Shakespeare did not write *Hamlet*. Consider the worlds $e(\sim s, w)$ and $c(\sim s, w)$. Because the second argument of the selection functions is w in each case, a comparison of their applications may suppress it. When applied with respect to the actual world, the selection functions reduce to one-place functions of propositions. Accordingly, given the actual world as background, $e(\sim s)$ and $c(\sim s)$ designate worlds obtained using, respectively, evidential and causal methods of filling out the supposition that Shakespeare did not write *Hamlet*. The evidentially nearest world in which Shakespeare did not write *Hamlet* is $e(\sim s)$. The causally nearest world in which Shakespeare did not write *Hamlet* is $c(\sim s)$. The proposition that another person wrote *Hamlet* is true in $e(\sim s)$ but false in $c(\sim s)$. Evidence preserves *Hamlet*'s existence, whereas causation does not.

Different types of supposition measure differently the distance between worlds and so yield selection functions that, given a proposition and the actual world as background, pick out different worlds as the nearest in which the proposition is true. The moods for an assertion of a conditional express different interpretations of distance between worlds. The indicative mood makes worlds that preserve evidential relations near to the actual world. The subjunctive mood makes worlds that preserve causal relations near to the actual world. The existence of *Hamlet* is good evidence that someone wrote it. So, among worlds in which Shakespeare does not write *Hamlet*, worlds in which someone else does are, evidentially speaking, closer to the actual world than worlds in which no one else does. On the other hand, the rarity of duplication of literary masterpieces makes *Hamlet*'s production by another author miraculous and contrary

to causal laws. So, among worlds in which Shakespeare does not write *Hamlet*, worlds in which no one else does are, causally speaking, closer to the actual world than worlds in which someone else does.¹¹

One may also make this point by contrasting ways of minimally revising the actual world to accommodate the supposition about Shakespeare. The supposition introduces a world in which Shakespeare did not write *Hamlet*. In such a world, other departures from the actual world are necessary. As a matter of logic, it cannot be that Shakespeare wrote *Hamlet* and *King Lear*. If causal laws are to be preserved, it cannot be that Shakespeare puts the words of *Hamlet* on paper, for that would be causally sufficient for his writing *Hamlet*. If evidential laws are to be preserved, it cannot be that *Hamlet* does not exist, for in that case innumerable readers are deceived by their senses. Revising the actual world to preserve causal laws removes *Hamlet* along with Shakespeare's authorship of *Hamlet*. Revising the actual world to preserve evidential laws retains *Hamlet* despite removal of Shakespeare's authorship of *Hamlet*.

Joint rationality is a type of conditional rationality. It involves supposition of a profile's realization. The type of supposition is evidential. One typically states the supposition in the indicative mood. Evidential supposition of a profile lets joint rationality characterize solutions to ideal games. Joint rationality assesses a strategy profile by considering whether each strategy is rational given that the profile is realized. When one considers whether if a profile is realized, an agent's strategy is rational, one considers whether the strategy is rational given the information the agent has if the profile is realized. The agent's evidence adjusts to accommodate the profile's realization. Background assumptions about an agent's knowledge may ensure that if a profile is realized, the agent is cognizant of its realization. In ideal games, evidentially supposing a profile's realization involves supposing knowledge of its realization. Realization of the profile replaces an agent's current evidence about other agents with new evidence about them. A profile's evidential supposition may attribute new beliefs to an agent even if the profile's realization has no causal influence on the agent's beliefs.¹²

Universal rationality is also a type of conditional rationality. It involves supposition of a profile's realization. The type of supposition is causal. One typically states the supposition using the subjunctive mood. Universal rationality assesses a strategy profile by considering whether, were the profile realized, all its strategies would be rational. Perhaps the profile's realization would cause an agent to have new beliefs. Then its supposition attributes new beliefs to the agent. In contrast, if the profile's realization would not causally influence the agent's beliefs, then its supposition preserves the agent's actual beliefs.

Joint and universal rationality diverge because they involve different types of supposition. A profile is universally rational if and only if, were it realized, all strategies in it *would* be rational. The supposition of the profile is causal. A profile is jointly rational if and only if, given its realization, all strategies in it *are* rational. The supposition of

¹¹ This illustration puts aside ties for nearest antecedent-world and the vagueness of the concept of distance between worlds. Resolving those issues is not crucial for my general points about types of supposition.

¹² The indicative supposition of a profile agrees with the indicative supposition of a strategy in an individual's strategic reasoning and so makes a solution amenable to support by individuals' strategic reasoning.

Fig. 3 Coordination

	Left	Right
Top	1, 1	0, 0
Bottom	0, 0	1, 1

the profile is evidential. The difference in a profile's causal and evidential supposition yields a difference in evaluation of its acts.¹³

Universal and joint rationality overlap in some cases. A profile realized has universally rational acts if and only if its acts are jointly rational. All ways of supposing a profile realized yield the same supposition and the same classification of its acts as rational or irrational. However, universal and joint rationality classify differently unrealized profiles. Both universal and joint rationality assess a profile by supposing its realization and then considering the rationality of acts in the profile. Universal rationality supposes the profile in a way that preserves causal relations, whereas joint rationality supposes the profile in a way that preserves evidential relations. A profile's causal supposition does not carry the same information that its evidential supposition carries. Rationality may classify an individual's act differently under a profile's causal supposition and under its evidential supposition.

Consider again the previous section's coordination problem, which Fig. 3 repeats. The game has two equivalent Nash equilibria in pure strategies, namely, (T, L) and (B, R) . Each equilibrium has jointly rational strategies given standard background assumptions about agents' knowledge. Under evidential supposition of an equilibrium profile, each strategy in it is rational. Take (T, L) . Given that Row performs T and Column performs L , it is rational for Row to perform T and for Column to perform L . The evidential supposition of the profile's realization carries information of the profile's realization. Given that the agents are aware of (T, L) 's realization, each strategy in the profile is rational. Similar points apply to the equilibrium profile (B, R) .

Suppose, as earlier, that a convention applies to the coordination problem. According to the convention, the players should realize (T, L) . Both players observe the convention, as is rational. The profile (T, L) is realized. It is jointly and universally rational. Consider the profile (B, R) . It achieves joint rationality for the reasons the previous paragraph reviewed. It need not achieve universal rationality, however. Suppose that the players cannot communicate to arrange a meeting at the nonconventional place. Then (B, R) does not achieve universal rationality. Causal supposition of the profile does not carry information of the profile's realization. No causal mechanism conveys information of its realization to the agents because they are incommunicado. Respecting causation, supposition of the profile leaves the agents in ignorance of its realization. If it were realized, both strategies in the profile would be irrational. Each would be performed in ignorance of the other. Neither Row nor Column would have any reason to perform the strategies the profile assigns.

Representing suppositions with selection functions highlights the contrast. Because the world the example describes provides a common background for suppositions,

¹³ Stalnaker (2005) shows the importance for game theory of the distinction between causal and evidential supposition. They have different roles in forward and backward induction, for instance.

one-place selection functions suffice. First, apply the evidential selection function to the profile (B, R) . In the world $e((B, R))$ Row believes that R is realized and so maximizes utility by performing B . Next, apply the causal selection function to the profile (B, R) . In the example's world, Column performs L , and Row believes that she does. If Column were to perform R instead, Row would still believe that she performs L . Row would not learn of Column's departure from convention. Hence, in the world $c((B, R))$ Row believes that L is realized and so does not maximize utility by performing B . Evidential and causal supposition of (B, R) differently affect Row's beliefs and so the strategy at the top of his utility ranking of strategies. That difference explains why the profile (B, R) achieves joint rationality but not universal rationality.

Considering minimal revision of the example's world presents the same explanation another way. If according to the evidential dynamics of a world w , a proposition p 's realization involves an agent's knowledge of p , then that knowledge is part of $e(p, w)$. If according to the causal dynamics of w , p 's realization does not yield an agent's knowledge of p , then that knowledge is not part of $c(p, w)$. Suppressing the constant background world, the agent knows that p in $e(p)$ but not in $c(p)$. According to the example's evidential dynamics, if the agents realize (B, R) , then they know that they realize this profile. Hence that knowledge is part of $e((B, R))$. According to the example's causal dynamics, if the agents realize (B, R) , then Row does not know that he and his partner realize this profile. He knows only that he realizes B . So knowledge of the profile's realization is not part of $c((B, R))$. The difference in Row's knowledge explains why B is rational given the profile's evidential supposition but is not rational given the profile's causal supposition.

What about moving from universal rationality to joint rationality? Does an inference in that direction work? No, a profile may achieve universal rationality without achieving joint rationality. Suppose that in a variant of the example if (B, L) were realized, it would be because Row believes incorrectly that Column goes to the unconventional meeting place and because Column believes that Row goes to the conventional meeting place. Then if (B, L) were realized, all its strategies would be rational. The profile achieves universal rationality. However, it does not achieve joint rationality. Given evidential supposition of the profile and the information that supposition carries, its strategies are not both rational. Because under the supposition Row learns that Column follows convention, he should also comply by performing T .

The difference between evidential and causal selection functions explains the difference between joint and universal rationality in this new version of the coordination problem. In the world $e((B, L))$ Row correctly believes that L is realized and so does not maximize utility by performing B . In the world $c((B, L))$ Row believes incorrectly that R is realized and so maximizes utility by performing B . Evidential but not causal supposition of the profile gives Row a true belief about Column's strategy. Evidential but not causal supposition of the profile makes its first component irrational.

Under special conditions, universal and joint rationality agree. In some ideal games, realization of a profile, either evidentially or causally supposed, yields knowledge of its realization. Communication may be perfect, for example. Then assessing a profile for universal rationality amounts to assessing it for joint rationality. Both assessments attribute to each agent the same beliefs and utility rankings of strategies. Also, in some nonideal games a profile's realization, either evidently or causally supposed,

does not affect agents' information and so the strategies rational for them. Agents may be completely unaware of their game and each other. In such games, universal and joint rationality also agree.¹⁴

Is joint rationality the same as collective rationality? They differ concerning profiles not realized, just as joint and universal rationality differ concerning those profiles. Take the coordination problem of Fig. 3. Under standard assumptions, if both agents were to depart from convention, they would each fail to be rational and so would together fail to be collectively rational. However, the profile in which each goes to the unconventional meeting place (B, R) is jointly rational because that profile's evidential supposition carries information of its realization. Also, if a failure to coordinate, such as (B, L), would arise from mistaken beliefs, as in the variant of the problem, then it may involve rational strategies and so achieve both universal and collective rationality. Nonetheless, a profile of uncoordinated strategies does not have jointly rational strategies.

Joint rationality and collective rationality agree in special cases, however. In ideal games, universal and joint rationality agree, and, in all games, universal rationality entails collective rationality. Hence, in ideal games, joint rationality entails collective rationality. Also, because in ideal games collective rationality entails universal rationality, collective rationality entails joint rationality in these games. Therefore, collective and joint rationality agree in ideal games.

How do joint and collective rationality compare with respect to attainability? Collective rationality is attainable in all games both ideal and nonideal. On the other hand, joint rationality is not attainable in every game. Different profiles yield different circumstances, and their suppositions affect differently a strategy's evaluation. For every profile, it may be that some agent's act is irrational given the profile. Take the pure-strategy game of Matching Pennies, which Fig. 2 presents. If supposition of a profile carries knowledge of the profile, no profile contains only strategies that maximize utility given the profile. Given that rationality requires utility maximization, no profile is jointly rational.¹⁵

4 Cooperative games

Game theory reveals paths to collective rationality and to goals of collective rationality. If the players in a game realize a subjective solution, they achieve joint rationality, and so universal rationality, and so collective rationality. Realizing an objective solution is a goal of collective rationality, and idealizations put the goal in reach. Its realization

¹⁴ Kadane and Larkey (1982) take a solution to be a profile of rational strategies. They argue that rational agents achieve universal rationality and not necessarily joint rationality. This point is correct. However, under the idealizations game theorists typically assume, agents anticipate the choices of others. Then joint rationality is necessary for universal rationality.

¹⁵ Weirich (1998, Chap. 5, Sect. 7.2) examines joint rationality's attainability. Standards of rationality do not demand utility maximization in every case. Hence, joint rationality is attainable in ideal games with a finite number of agents even when the games lack Nash equilibria. However, joint rationality is unattainable in some nonideal finite games.

does not require that players intend to achieve it. Each may pursue private goals. In ideal conditions that behavior nonetheless realizes an objective solution.

Cooperative games have features that promote attainment of goals of collective rationality. They permit communication and binding contracts. They offer opportunities for joint action. Pareto optimality is a goal of collective rationality. Which ideal conditions, besides standard ones about players' cognitive abilities and comprehension of their game, ensure that universal rationality brings the group to that goal?

In cooperative games agents may act jointly. That is, they may causally coordinate their acts using communication and binding agreements. Rationality may require agents to use their opportunities for joint action. Because cooperative games afford opportunities for joint action, it may seem that in them individual rationality leads to Pareto optimality. Indeed, transforming the Prisoner's Dilemma into a cooperative game makes the players' rationality yield Pareto optimality. However, rational individuals may fall short of Pareto optimality in other cooperative games.

Consider the Ultimatum Game. It is a two-person bargaining problem. The two players have \$10 to divide if they agree on a division. The first player makes a proposal, and the second may only take it or leave it. If the second player declines, neither player gains anything. Suppose that the two players are involved in a series of repetitions of the game. In a play of the game, the first player proposes \$9 for himself and \$1 for the second player. The second player rejects the low offer to establish a reputation for toughness. The first player's low offer is rational if he does not anticipate the second player's intention to build a reputation, and the second player's refusal is rational given its anticipated benefit in future games. Because the players achieve universal rationality, they achieve collective rationality despite failing to achieve Pareto optimality. Although players in a cooperative game have opportunities for joint action, their communication need not be perfect. In nonideal circumstances rational players may misunderstand each other and as a result fail to achieve Pareto optimality.

Consider a cooperative game under the idealization of perfect communication. Does that idealization suffice for Pareto optimality? In a simple coalitional game, a characteristic function specifies the value of each coalition, that is, the value of the coalition's best joint action. In the game, joint rationality evaluates coalitions as well as individuals. A strategy profile is jointly rational if each component, including a coalition's joint act, is rational given the profile. Under standard assumptions, the core is the set of profiles that achieve joint rationality.

Rational agents in a coalitional game with perfect communication may fail to achieve Pareto optimality. Figure 4 presents a simple game with three persons *A*, *B*, and *C*. Combinations of the letters *A*, *B*, and *C* stand for coalitions, and *v* is a characteristic function specifying the value of each coalition.

Imagine that conditions for communication and joint action are perfect. Because the game's core is empty, the agents cannot realize a core allocation. They may

Fig. 4 Exclusion

$$\begin{aligned} v(A) &= v(B) = v(C) = 0 \\ v(AB) &= v(BC) = v(AC) = 8 \\ v(ABC) &= 9 \end{aligned}$$

form the coalition structure $\{\{A, B\}, \{C\}\}$ and achieve the outcome profile $(4, 4, 0)$. That outcome achieves universal rationality and so collective rationality. However, it excludes a productive member of the group and is not Pareto optimal.

Ideal conditions besides perfect communication must be introduced to ensure attainment of Pareto optimality in cooperative games. Adding the idealization that agents are comprehensively rational does the job. Their comprehensive rationality includes their adequately preparing for coalitional games and so regulating pursuit of incentives to achieve Pareto optimality. Rational preparation for the game seizes all opportunities for productivity.

To support this point about preparation for a cooperative game, consider first coordination problems, such as the one Fig. 3 depicts. Treatments of coordination distinguish rational and hyperrational agents. According to the prevalent account, a hyperrational agent does his part to achieve coordination only if convinced others will do their parts. For agents with common knowledge of their hyperrationality, coordination fails. Nothing gets the ball rolling. However, rational agents have more freedom than hyperrational agents do. Rational agents may take steps to initiate coordination. For example, they may create conventions to foster coordination, conventions such as driving on the right side of the road. Rationality does not impose hyperrationality's constraints.

This point about rationality applies also in coalitional games. Rational preparation for a coalitional game yields coordination for productivity. Comprehensively rational agents prepare for the game. If conditions are ideal for joint action and agents are ideal and comprehensively rational, they achieve Pareto optimality.

A full theory of collective rationality elaborates this brief account of attainment of goals of collective rationality. It explains in detail the preparations rational agents make to achieve Pareto optimality. Because this paper's objective is just to introduce collective rationality, and not to present a full theory of collective rationality, it does not undertake that elaboration.

5 A general theory of rationality

Collective, universal, and joint rationality all have roles in a general theory of rationality. Collective rationality is rationality's extension to groups. Universal rationality is an individualist's recommendation to groups. Joint rationality is game theory's characterization of a solution. The three types of evaluation resemble each other but are distinct.

Rationality requires a group of people to be collectively rational. It does not require a group to achieve universal or joint rationality, except when those types of rationality agree with collective rationality. Neither universal nor joint rationality agrees with collective rationality in all cases. They reliably reveal rationality's standards for groups only in ideal cases.¹⁶

¹⁶ For valuable comments, I am indebted to Maurice Salles and participants at the Conference on the Philosophical Aspects of Social Choice Theory and Welfare Economics held at the University of Caen, June 20–21, 2005.

References

- Arrow K (1951) Social choice and individual values. Yale University Press, New Haven
- Aumann R (1987) Correlated equilibrium as an expression of Bayesian rationality. *Econometrica* 55:1–18
- Bernheim BD (1984) Rationalizable strategic behavior. *Econometrica* 52:1007–1028
- Broome J (1987) Utilitarianism and expected utility. *J Philo* 84:405–422
- Joyce J (1999) The foundations of causal decision theory. Cambridge University Press, Cambridge
- Kadane J, Larkey P (1982) Subjective probability and the theory of games. *Manage Sci* 28:113–120
- Nash J (1950) Equilibrium points in N-person games. *Proc Nat Acad Sci* 36:48–49
- Pearce D (1984) Rationalizable strategic behavior and the problem of perfection. *Econometrica* 52:1029–1050
- Pettit P (2001) A theory of freedom: from the psychology to the politics of agency. Oxford University Press, New York
- Rescher N (1988) Rationality: a philosophical inquiry into the nature and the rationale of reason. Clarendon Press, Oxford
- Sen A (1977) Rational fools. *Philo Public Aff* 6:317–344
- Sen A (2002) Rationality and freedom. Harvard University Press, Cambridge
- Stalnaker R (2005) Counterfactual propositions in games. Manuscript presented at the 2005 Pacific Division APA meeting in San Francisco
- von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton
- Weirich P (1998) Equilibrium and rationality: game theory revised by decision rules. Cambridge University Press, Cambridge
- Weirich P (2004) Realistic decision theory: rules for nonideal agents in nonideal circumstances. Oxford University Press, New York