Schooler 1991; Anderson et al. 1997; Flood 1954; Jones & Sieck, in press; Oaksford & Chater 1996; Schacter 1999).

The difficulty with the CKR axioms is that they require players to reason about their opponents entirely a priori, based only on the assumptions of rationality and common knowledge, while ignoring all other potential sources of information. A more faithful model of rational choice would allow players to utilize all the knowledge available to them, including general knowledge about human behavior or specific knowledge about the opponent gained from previous interactions (e.g., earlier moves). For example, the fact that the conventional priority of Heads over Tails leads to the phenomenon of focal point selection should realistically be available to each player as information for use in predicting the opponent's choice. Thus, all that is needed is a simple intuitive understanding of human behavior for a subject to infer correctly (and rationally) that the opponent is likely to choose the focal option. Instrumental rationality then dictates that the player chooses that option as well. Similar reasoning applies to payoff dominance in the case of the Hi-Lo matching game.

Relaxing the restrictions provided by CKR on players' models of their opponents can also explain violations of the prescriptions of backward induction arguments. If Player II's model of Player I admits alternatives to perfect rationality, then an initial cooperative move by Player I will simply lead to an update of II's beliefs about I (rather than generating a logical impasse). This sort of updating can be formalized using a Bayesian framework, in which each player has probabilistic prior beliefs about the opponent (perhaps peaked around rationality, but nonzero elsewhere), which are determined by prior experience with the opponent or with people in general. Even if the prior expectation were heavily biased towards strict rationality, an initial cooperative move by Player I would force Player II's model to favor other possibilities, for example, that Player I always plays Tit-For-Tat. This could lead to Player II cooperating on step 2, in turn giving Player I justification for cooperating on step 1.

The preceding arguments have shown how failures of CKR can be remedied by more complete normative analyses that preserve the assumption of instrumental rationality, that is, optimality of actions as conditioned on the model of the opponent. The question of rationality in game scenarios then shifts to the rationality of that model itself (inductive rationality). In the case of focal point selection, we have offered no specific mechanism for the inductive inference regarding the opponent's likely choice, as based on general experience with human behavior. We merely point out that it is perfectly consistent with the assumption of inductive rationality (although it has no basis in CKR). (Ironically, the same empirical fact that is cited as evidence against RCT – namely, focal point selection – actually corroborates the rationality of people's inductive inferences.)

The stance taken in our discussion of backward induction, whereby people are rational yet they entertain the possibility that others are not, presents a subtler problem. What must be remembered here is that, as a positive theory, RCT only claims that people try to act rationally (target article, sect. 3.3), and that the idealization of perfect rationality should give qualitatively correct predictions. Of course, in reality, people do err, and subjects are aware of this fact. Therefore, in forming expectations about their opponents' actions, subjects are open to the possibility of errors of reasoning by the opponent. Furthermore, as one progresses further back in the chain of reasoning entailed by backward induction, the expectation of such errors compounds. Thus, the framework proposed here can be viewed as idealizing rationality at the zero level, but not at higher orders of theory-of-mind reasoning.

Our thesis, that people follow instrumental rationality but anchor it on their model of the opponent, is supported by Hedden and Zhang's (2002) recent investigation of the order of theory-of-mind reasoning employed by subjects in three-step sequential-move games. On each trial, subjects, who controlled the first and third moves, were asked first to predict the response of the opponent (a confederate who controlled the second move) and their own best choice on the first move. Initially, subjects tended to predict myopic choices by the opponent, corresponding to level 0 reasoning (level 1 was optimal for the opponent). Accordingly, subjects' own actions corresponded to the level 1 strategy, rather than the level 2 strategy prescribed by CKR. However, after sufficient experience with an opponent who played optimally, 43% of subjects came to consistently predict the opponent's action correctly, and altered their own behavior to the level 2 strategy. Although the remaining subjects failed to completely update their mental model of the opponent, errors of instrumental rationality (discrepancies between the action chosen and that dictated by the expectation of the opponent's response) remained low and approximately constant throughout the experiment for both groups. These results support the claim that violations of the predictions of CKR can be explained through scrutiny of player's models of their opponents, without rejecting instrumental rationality, and suggest that further investigations of rational choice in game situations must take into account the distinction between instrumental and inductive rationality.

Analogy in decision-making, social interaction, and emergent rationality

Boicho Kokinov

Central and East European Center for Cognitive Science, Department of Cognitive Science and Psychology, New Bulgarian University, Sofia, 1618 Bulgaria. **bkokinov@nbu.bg**
http://www.nbu.bg/cogs/personal/kokinov

Abstract: Colman's reformulation of rational theory is challenged in two ways. Analogy-making is suggested as a possible candidate for an underlying and unifying cognitive mechanism of decision-making, one which can explain some of the paradoxes of rationality. A broader framework is proposed in which rationality is considered as an emerging property of analogy-based behavior.

Rationality has long been shown to fail as a descriptive theory of human decision-making, both at the individual and social levels. In addition, Colman presents strong arguments that rationality also fails as a normative theory for "good" decision-making – "rational" thinking does not produce optimal behavior in social interaction and even acts against the interests of the individual in some cases. Fortunately, human beings often act against the postulates of rationality and achieve better results than prescribed by the theory. Therefore, Colman concludes that "rationality" has to be redefined by extending it with additional criteria for optimization, such as the requirement for maximizing the "collective" payoff, or with additional beliefs about the expected strategies of the coplayers. He does not clarify how and when these additional criteria are triggered or where the common beliefs come from.

We are so much attached to the notion of rationality that we are always ready to repair it, but not to abandon it. The theory of rationality is, in fact, a formalization of a naive theory of human thinking. This naive theory makes it possible to predict human behavior in most everyday situations in the same way as naive physics makes it possible to predict natural phenomena in everyday life. However, no one takes naive physics so seriously as to claim that it provides "the explanation" of the world. Moreover, even refined and formalized versions of this naive theory, like Newtonian mechanics, are shown not to be valid; and more complicated and counterintuitive theories at the microlevel, like quantum mechanics, have been invented. On the contrary, rationality theory is taken seriously, especially in economics, as an explanation of human behavior.

Instead of extending rationality theory with additional socially oriented rules, it may be more useful to make an attempt to build a multilevel theory that will reveal the implicit and explicit cognitive processes involved in decision-making. These underlying cog-

https://doi.org/10.1017/S0140525X03370050 Published online by Cambridge University Press

nitive mechanisms produce decisions, which are sometimes "individually rational," sometimes "collectively rational," and sometimes "not rational at all." Because these mechanisms have been evolved and developed to assure human survival, they will, most of the time, produce results that are "rational" or "optimal" from some point of view – this is what makes rationality a good naive theory. However, this does not mean that people explicitly follow the rules of maximization prescribed by the theory.

Colman proposes an eclectic collection of ad-hoc strategies (team reasoning, Stackelberg reasoning, epistemic, and nonmonotonic reasoning), which are all different forms of explicit deductive reasoning. Deduction can certainly play a role in decision-making, but it is not enough to explain it. Recent studies revealed that analogy-making is a more basic mechanism of human thinking, which is present from early infancy and is used ubiquitously in everyday life (Gentner et al. 2001). Analogy-making is a process of perceiving one situation (target) in terms of another (base), thereby preserving the system of relations among elements and transferring knowledge from the base to the target. Arguments have been presented that deduction is in fact based on analogy, and a special form of it (Halford 1993; Kokinov 1992). Markman and Moreau (2001) have reviewed the evidence that analogy plays an important role in perceiving and framing the decision situation, as well as in comparison of the alternatives. Moreover, analogy may be used both explicitly and implicitly (Kokinov & Petrov 2001; Markman & Moreau 2001). Thus, analogy may play a unifying role in describing the mechanisms of decision-making.

Analogy-making may explain the paradoxes of using the focal points described by Colman. They are easily perceivable and analogous to focal points in other games. Therefore, it is natural to expect people to use them again and again if previous experience of using a focal point has been successful. Similar arguments may be applied to social dilemmas and trust games. If another player has used a certain strategy in a previous case, I may expect him or her to behave the same way in an analogous situation, and thus have a prediction for his or her behavior.

Analogies may be applied at various levels: Analogies to previous cases of decision-making in the same game or analogies to games with similar structure; analogies to cases of social interaction with the same individual or to cases of social interactions with individuals who are considered analogous (i.e., are in similar relations to me, like family or team members). Thus, even a novice in a particular game can still use his or her previous experience with other games.

Analogy can explain the "deviations" from the prescribed "rational" behavior and the individual differences among players. If a player has an extensive positive experience of cooperative behavior (i.e., many successful cases of benefiting from acting together), and if the current game is found to be analogous to one of these cases, then he or she might be expected to act cooperatively (even if this is not the optimal strategy). On the contrary, if the game reminds the player of a previous case of betrayal or fraud, then defection strategy should be expected.

In summary, analogy may play a crucial role in a future theory of decision-making. Instead of explaining rationality with rules for utility maximization, which people follow or break, we may explain human behavior by assuming that decisions are made by analogy with previous cases (avoid strategies that were unsuccessful in analogous situations and re-use strategies that were successful). Thus, utility maximization is an emergent property that will emerge in most cases, but not always. In this view, rationality is an emergent phenomenon, and rational rules are only a rough and approximate explanation of human behavior.

Wanted: A reconciliation of rationality with determinism

Joachim I. Krueger

Department of Psychology, Brown University, Providence, RI 02912.
joachim_krueger@brown.edu
http://www.brown.edu/departments/psychology/faculty/krueger.html

Abstract: In social dilemmas, expectations of reciprocity can lead to fully determined cooperation concurrent with the illusion of choice. The choice of the dominant alternative (i.e., defection) may be construed as being free and rational, but only at the cost of being incompatible with a behavioral science claiming to be deterministic.

The conspicuous failure of orthodox game theory is its inability to account for cooperative behavior in noniterated social dilemmas. Colman outlines a psychological revision of game theory to enhance the predictability of hitherto anomalous behavior. He presents the Stackelberg heuristic as a form of evidential reasoning. As Colman notes, evidential reasoning is assumed to lead respondents to shun the dominating alternative in Newcomb's problem and in decisions to vote. In the prisoner's dilemma game (PDG), however, Stackelberg reasoning leads to defection (Colman & Stirk 1998). Thus, Stackelberg reasoning appears to be neither evidential nor parsimonious in this domain. After all, players can select the dominating alternative in the PDG without making any predictions of what their opponents will do. How, then, can evidential reasoning lead to cooperation?

The logic of the PDG is the same as the logic of Newcomb's problem (Lewis 1979). Just as players may expect that their choices will have been predicted by Newcomb's savvy demon, they may expect that their choices in the PDG will most likely be matched by their opponent's choices (unless the rate of cooperation is exactly 50%). The issue is whether this statistical realization gives cooperators (or one-boxers, in Newcomb's case) license to lay claim to being rational.

Orthodox game theorists insist on defection, because a player's cooperation cannot make an opponent's cooperation more likely. Evidentialists, however, claim that cooperation may be chosen without assuming a causal effect on the opponent's choice. Only the assumption of conditional dependence is needed. If nothing is known about the opponent's choice, conditional dependence is obvious *after* a player committed to a choice. By definition, most players choose the more probable alternative, which means that the choices of two independent players are more likely to be the same than different (Krueger 1998). Because time is irrelevant, it follows that it is more likely that two players *will* make the same, instead of different, choices. In the extreme case, that players expect their responses to be reciprocated without fail, their dilemma devolves into a choice between mutual cooperation and mutual defection. As mutual cooperation offers the higher payoff, they may choose cooperation out of self-interest alone.

Evidentialist reasoning is distasteful to the orthodox mind because it generates two divergent conditional probabilities that cannot both be correct (i.e., p[opponent cooperation/own cooperation] $>$ p[opponent cooperation/own defection]). Choosing the behavior that is associated with the more favorable prospect then smacks of magical thinking. But causal assumptions enter at two levels: at the level of the investigator and at the level of the participant. Investigators can safely assume that players' efforts to influence their opponents are pointless. Players, however, may *think* they can exert such influence. Although this expectation is irrational, it does not invalidate their cooperative choices. Note that investigators can also subscribe to a more plausible causal argument, which holds that both players' choices result from the same set of latent variables. These variables, whatever they may be, produce the proportions of cooperation found in empirical studies. Players who realize that one option is more popular than the other, but do not know which, can *discover* the popular choice by observing their own. The fact that they may have an experience of

https://doi.org/10.1017/S0140525X03370050 Published online by Cambridge University Press