Studies in Logic, Vol. 17, No. 3 (2024): 102–128
PII: 1674-3202(2024)-03-0102-27

The Independence Thesis in Practice: How to Explain Group Epistemic Success*

Patricia Rich Hailin Liu✉

Abstract. An important lesson from social epistemology is the two-part "independence thesis": groups may be irrational despite each member being rational, and groups may be rational although their members are irrational. We urge skepticism regarding real-world analyses of the second type, using the research on rationalizing cooperation as a paradigmatic example. We discuss a prominent example of such analysis—information cascades—and argue that it can be fruitfully re-conceptualized as an instance of epistemic cooperation. This new perspective points the way towards improved explanation and understanding of this important scenario.

1 Introduction

An important lesson from social epistemology is the so-called "independence thesis", which tells us that individual and group rationality are theoretically independent ([45, 46]). Nor do they seem to (nonetheless) coincide in practice. There are familiar examples in which individual rationality does not ensure group rationality. It is also apparently common for members of a group to behave in a way that is individually epistemically irrational, but for the whole group to be more epistemically successful—e.g. find the truth more quickly or reliably—as a result.

It is one thing to recognize the theoretical independence between individual and group rationality, and quite another thing to identify real-world phenomena as reflecting this independence. In particular, we argue for an explanatory asymmetry between the two parts of the independence thesis. One part of the thesis is that groups may be irrational despite consisting of rational individual members. Given that this is possible, we see no reason not to expect the problem to arise for real-world groups. It may

Received 2023-04-06 *Revision Received* 2023-09-11

Patricia Rich Department of Philosophy, University of Bayreuth
patricia.rich@uni-bayreuth.de
Hailin Liu✉ Institute of Logic and Cognition, Sun Yat-sen University
Department of Philosophy, Sun Yat-sen University
liuhlin3@mail.sysu.edu.cn

*Patricia Rich was supported by ERC project "Epistemic Utility Theory: Foundations and Applications" (No. ERC-2012-StG_20111124) and the DFG project "Knowledge and Decision" (No. 315078566) while this work was carried out. Hailin Liu acknowledges the financial support received from the Major Program of National Social Science Foundation of China (No. 23&ZD240). The authors would also like to thank audiences in Bristol, Guangzhou, Glasgow, Hamburg, and Ischia for discussion. Dominik Klein, Samir Okasha, and several referees provided helpful, detailed comments.

be unfortunate that individual rationality fails to coordinate for the group good, but an explanation of this form raises no red flags. The other part of the thesis is that a group may be rational despite having individually-irrational members. Although this is possible in principle, we should be skeptical of explanations of real-world phenomena having this form, and especially skeptical of the idea that groups *commonly* owe their success to the systematic irrationality of their members.

The present paper both elaborates on this explanatory asymmetry and demonstrates that it has important implications for research in social science and social epistemology, where it has been neglected in some high-profile cases. To this end, we discuss a prominent real-world phenomenon which has been analyzed in the literature as a case of group rationality depending on individual irrationality. We explain how the traditional explanations are problematic, and show that the initial verdict of individual irrationality was too hasty, depending on a narrow selection of theoretical models. We show that it is not difficult to find competing rationalizing explanations of the individual behavior in question, once one recognizes that these must be sought.

The phenomenon that this paper discusses is *information cascades*. In an information cascade, an entire group converges to a weakly-supported opinion because public reports outweigh private evidence. For example, a restaurant may become much more popular than another among users of a particular ratings website solely because a few bogus reviews cause people to try it first. Individually-rational conformity is harmful to the group as a whole, making it less reliable. Empirically, however, we are not all conformists; some diners will ignore the many glowing reviews and follow their neighbor's alternative recommendation instead. Sticking to one's own view is often taken to be irrational, but is not uncommon, and it benefits the group by injecting more independent evidence into the public pool. We show that this problem can be re-framed as a Prisoner's Dilemma, and characterize sticking to one's private evidence in a cascade situation as an epistemic form of cooperation. We then explain how information cascade problems can be fruitfully re-analyzed by translating the plethora of strategies for rationalizing individual cooperation in the Prisoner's Dilemma literature to the epistemic domain.

This paper is organized as follows: Section 2 argues for the explanatory asymmetry between the two parts of the independence thesis and explains its importance. Section 3 discusses information cascades, and shows how to use the Prisoner's Dilemma literature to provide alternative, rationalizing explanations of non-cascading behavior. Section 4 concludes the paper with a brief discussion.

2 Explaining Group Rationality

In this section, we defend the explanatory asymmetry between the two parts of the independence thesis. We first explain how group rationality and individual rationality are (and should be) treated differently. However, explanations of group rationality as the result of individual irrationality neglect such a difference. We then show how this supports the explanatory asymmetry that we have identified. Finally, we briefly consider the process of explaining human cooperation in the practical domain, as it unfolded in the biology and economics literature. This shows how the methodological principles we appeal to essentially drove the explanatory process in a setting analogous to ours.

For the purposes of this paper, "rationality" can be defined as utility maximization, broadly construed. Individual rationality therefore means that the individual acts so as to maximize their expected utility; in the epistemic domain, this implies that they update their beliefs based on their total evidence, for example, so that their beliefs are expected to be as accurate as possible. Group rationality is the analog as applies to a group. Exactly what it means for a group to maximize utility will depend on the context; it could mean, for example, that as many group members as possible have accurate beliefs or take the best action. While the independence thesis says that we can have individual rationality but group irrationality as well as the reverse ([45]), we should expect the former to be much more common. Group rationality and individual rationality are typically treated differently, and for good reason. Let us consider them individually, before considering how they are likely to combine.

Group rationality may take somewhat different forms in different settings, as noted above. However defined, it is a fact of social science methodology that we do not generally expect groups to be rational. Indeed, there is no particular reason why we should expect groups to be rational: Insofar as group performance results from the many independent actions of individuals, it would be quite surprising if these actions are always or even typically coordinated for the good of the group.

Examples to the contrary are ubiquitous and familiar. In the practical domain, game theory has uncovered and studied numerous social dilemma (see e.g. [19]). The best known is the Prisoner's Dilemma, which will be discussed in some detail in Section 3. In such situations, there are socially desirable outcomes which cannot be achieved when individuals choose rationally according to the theory; collective action would be beneficial, but is not possible. These social dilemmas have counterparts in biology; Frank ([23]) provides an interesting (though overtly political) discussion of the practical manifestations and implications of the point that individual rationality often harms the group greatly. Frank discusses cases involving both humans and other animals, and the most dramatic examples come from "arms races" in the animal kingdom. For example, the antlers of male bull elk have gotten larger and larger, since larger antlers make each individual male better able to compete for access to females. Nevertheless, the elk would be much better off as a group if their antlers were all, say, half as large. The antlers are now so large that the males cannot travel through the forest very easily and have trouble escaping from predators. The consequences of such "arms races" are unfortunate and can be devastating for species, but are not at all surprising given the logic of natural selection.

A theme underlying this paper is that epistemic and practical problems are analogous, and there is no reason to expect group epistemic rationality to differ from group practical rationality here. In fact, we know that it does not. We find similar examples of "irrational" epistemic groups, e.g., the Doctrinal Paradox. This paradox instantiates the common pattern in which natural ways of aggregating rational individual states produce an irrational group state, in this case the acceptance of a logically inconsistent set of propositions (see [42] for discussion). In these epistemic cases, as in practical cases, there is no particular reason to expect a group to be rational.

Individual rationality is treated differently, as a matter of methodological principle, which is often taken to be expected utility maximization based on all the available evidence. Specifically, we seek to model and explain individual behavior on the premise that it is somehow rational (a practice sometimes referred to as "methodological rationalism"). The social sciences have been characterized by their goal of interpreting human activities, in contrast with the natural sciences; this requires seeing humans as rational agents in some sense (see [18, 72] and [41] for discussion of their views). Methodological rationalism is most explicit in economics, where agents are nearly universally modeled as maximizing some kind of utility function (see [26] for a defense of orthodox methodology). Biologists extend this methodological principle to a much wider class of biological entities, which are similarly modeled as maximizing a utility function, with utility corresponding to fitness.[1]

While methodological rationalism could be taken to require modeling individuals as rational agents in a very specific sense (for example as Bayesian reasoners and expected utility maximizers in economics), this interpretation is neither necessary nor ours. Instead, the "rationality" involved has a much broader meaning. Lee ([41]) describes a methodological rationalism which is *Galilean* in the sense of Cartwright ([15]) and McMullin ([48]); rationality assumptions provide a way to systematize observations and generate specific hypotheses, but should be de-idealized as the science matures. She argues that this approach plays a key role in modern psychology, a role which is perhaps obscured by the fact that "rationality" is defined differently by different research programs. We agree with Lee, and our point is that we should try to explain human behavior as rational in some sense, rather than in a particular sense. The definition of rationality needs to be loosened precisely in order to maintain the premise of individual rationality. This is done not just for its own sake or because of a dogmatic conviction that people are rational, but because it is our best (and often our only) tool for understanding the determinants of or reasons behind human behavior. The discussion in the rest of the paper will help to illustrate this point.

It is critical to note that we are *not* arguing that people are in fact always rational (nor does methodological rationalism imply this). Our point is not about the status of individuals as rational or irrational, but about the process by which we go about

[1] See [53] for discussion; in contrast with [72], Okasha characterizes biology as a sense-making enterprise along with the social sciences (Ch. 2.4).

explaining human behavior. This process typically involves hypothesizing that the given behavior is rational in some particular sense, and then testing the hypothesis. If the hypothesis fails, the process is repeated. Now, even the most panglossian (perhaps someone like Gigerenzer, see [24]) would not argue that we are always perfectly rational, and the nature and extent of human rationality are the subject of intense, long-running debate (see [9, 38, 56, 61, 64, 66, 67, 70] for broad perspectives). Yet we observe that apparent instances of irrationality according to classical norms are seldom simply accepted at face value by the research community, and our understanding of human rationality would be extremely impoverished if they were.

Importantly, while there is broad agreement that there are instances of (even systematic) human irrationality, there is essentially no agreement regarding which instances they are. Consider some of the best-known contenders: The Allais and Ellsberg paradoxes reveal violations of Expected Utility Theory ([3, 21]), but there is disagreement regarding whether one or both reveals the irrationality of the deciders rather than limitations of the theory; rationalizing models dominate within economics, while philosophers' views are mixed (see e.g. [14, 22, 37, 49]). The Linda problem is taken by some to show that we are subject to the conjunction fallacy, and by others to show that rational subjects draw pragmatic or linguistic inferences which are not captured by logical models ([34, 69]). The situation is similar for the Wason Selection Task and modus ponens performance ([17, 51, 65, 71]).

One can of course debate whether we should dub individuals "rational" in each case, but the conclusion of particular debates is beside the point. We learn a great deal about human rationality through the debate process itself, in which a plethora of rationalization strategies are investigated. There is an important sense in which not engaging in this process amounts to giving up the project of explaining or understanding human choice and inference. Any given human behavior could be irrational, but we are not licensed to conclude this as a general, model-independent fact until we have analyzed the behavior from various angles, considered the context, and found that the best of our explanations is such that the behavior does not count as rational in a meaningful sense. Most often, we find that there is a sense in which the behavior is rational, and a sense in which it is not; together, they provide a robust understanding of the phenomenon. When it comes to rationalizing explanations of individual behavior, the situation is aptly described as "seek and ye shall find."

Let us now consider how group and individual rationality and irrationality are likely to combine, with an eye towards the independence thesis. Individual rationality is a standard and productive methodological premise, while group rationality is not. Therefore, explanations in which individuals are rational but groups are not are unexceptional. They raise no red flags. The opposite kind of examples have recently received a lot of attention, however. These are examples in which individuals supposedly behave or reason irrationally, but in doing so, they promote the rationality of the group. We should be doubly skeptical of such accounts.

Firstly, in the case of information cascades, it has been accepted that many individuals behave irrationally without the usual skepticism and extensive debate. It would be profitable, here as elsewhere, to work harder to explain superficially irrational behavior, whether it is ultimately fully rationalized or merely excused. This is just the application of the methodological principle described above. Individuals could be irrational, but we will not be able to explain their behavior satisfactorily until we consider a variety of alternative rationalizing hypotheses, and see how they fare.

Secondly, if the individuals in question were truly irrational, it would be quite a coincidence for their errors to systematically benefit the group as a whole—in not just one, but potentially many facets of our social-epistemic lives. More plausible explanations of group success would instead explain why the individuals do what benefits the group. For example, the group good may be sufficiently aligned with the individual good for the individually-rational behaviors to be those which are helpful, or at least not highly detrimental, to the group. So in cases in which group success is attributed to rampant individual error, there is even more cause than usual to investigate alternative possibilities. There is an additional justification for methodological rationalism in cases of apparent group rationality.

A comparison with practical rationality is instructive. Consider the project of explaining the widespread cooperation among humans, for example the fact that most of us do not harm or steal from one another even when we would intuitively benefit from doing so and could get away with it.[2] This is a case in which, at first glance, individual irrationality allows the group to succeed. Since this pattern of explanation was recognized as problematic, better and better rationalizing explanations have been explored. One proffered explanation is that cooperation with relatives promotes inclusive fitness, but this cooperation is implemented via a simple cooperative choice heuristic which we erroneously apply when interacting with non-relatives as well. Yet this is not biologically plausible: it essentially explains the long-term functioning of human society through constant mistakes on all of our parts. A second possibility is to explain cooperation through group selection: cooperative groups are more successful than non-cooperative ones. While this explanation does gesture at a reason for cooperation, it is still problematic because it does not explain how individual cooperation persists. So long as each individual can gain an advantage by harming and stealing, we should expect this to happen, even when this is detrimental to the group.[3] Group selection can be part of the explanation—as in Bowles and Gintis' account ([12])—but the explanation must also satisfyingly account for individual be-

[2]Birch provides a useful overview in [11].

[3]We paint this evolutionary picture with broad brush strokes. For an overview of the history of "group-selection" explanations, including present points of contention, see [52, Ch. 6]. See also [53, Ch. 2.3-4] for more practical discussions of which kinds of explanations are legitimate and appropriate in what circumstances, and Ch. 7.2 of the same book for a discussion of the rationality of evolved cooperative behavior in the case of the Prisoner's Dilemma.

havior. We seek an ultimate-level explanation of cooperation—an explanation of why the behavior exists—and this explanation must specify how it can be fitness maximizing for individuals to behave in such a way that the group is better off ([62]). Many such explanations have now been developed; this is methodological rationalism and our explanatory asymmetry at work.

Furthermore, the purportedly irrational epistemic behaviors under consideration are naturally construed as cooperative: they benefit the epistemic community. As in the practical case, then, it is crucial that we explore how cooperation could be rational for the individuals. We explain how the existing tradition of research into cooperation can be fruitfully translated into the epistemic domain, providing myriad hypotheses regarding how epistemic cooperation could be individually rational. To be clear, we do not deny that existing models can provide insight, and we do not directly criticize these models. Nonetheless, we must ask which factors, left out of existing models, require consideration for a fuller understanding. We must investigate which factors permit purportedly irrational behavior to be rationalized, as discussed in what follows.

3 Strategy in Information Cascades

3.1 The standard analysis

Crowd-following or herding behaviors are quite common. An information cascade is a particular type of herding event that can occur when individuals act in a sequence and only their actions are observed by subsequent actors; the individuals' private evidence pertaining to the best action is not observed by others. The private evidence of an individual may be outweighed by their public evidence, i.e. the actions others have already taken. As a result, the individual imitates the earlier actors and their private evidence has no impact. At the group level, this can lead to the entire group acting in the same way, and furthermore acting on the basis of just the few pieces of evidence possessed by the first individuals to act.[4]

Information cascades are troubling because the group can very easily converge on the wrong action, even though the individuals collectively have enough information to easily identify the correct action ([7]). Worse still, information cascades are not a mere theoretical possibility; they are thought to be very common in real life, in situations ranging from the benign (e.g. restaurant and book choices) to the momentous (stock market bubbles and crashes [33] and political watersheds [43]. See [32, Ch. 4] for many real-life examples).

It is clearly important that we can correctly analyze these situations, so that we can better predict when information cascades will occur and how strong or robust they will be. In the most weighty situations (such as political elections or economic

[4] One of the authors also discusses information cascades in [57].

downturns), it is desirable to know what possibilities exist for preventing information cascades. Understanding why and how they occur (or do not) is a critical part of this. Yet information cascades are a prime example of the explanatory asymmetry between the two sides of the independence thesis being neglected. The received view is that rational individuals cascade, making groups unreliable. As we explain here, this account is far from empirically adequate and normatively satisfactory. In real-world and experimental settings, many individuals do not cascade; according to the received view, then, groups are made more reliable by these individuals' irrationality. But this account is theoretically problematic for the reasons explained in Section 2; the non-cascading behavior remains a mysterious source of good fortune. This section discusses information cascades in detail, showing how acknowledging and responding to the explanatory asymmetry provides us with a vast new toolbox for understanding these situations.

The key features of information cascades are best illustrated through an example. This example is initially due to Bikhchandani et al. ([10]) and was first studied in a laboratory setting by Anderson and Holt ([4]). A group is shown with two urns, one in which two-thirds of the balls are black and one-third are white, and the other in which the reverse is true. One urn is then chosen at random, and the group knows that both are equally likely. Participants take turns receiving a private signal about the urn (seeing the color of one ball that is drawn at random, with replacement) and making a public guess as to which urn has been chosen. Participants are paid for correct guesses and receive nothing for incorrect guesses. This procedure is repeated, often for 15 rounds.

An information cascade can develop as follows: Suppose the first two participants draw white balls, and publicly guess that the urn is the majority white urn. Next, suppose that the third participant draws a black ball. Although her private evidence supports the majority black urn, the majority white urn is more likely on her total evidence, which is that (presumably) two white balls and one black have been drawn. The third participant therefore guesses that the majority white urn has been chosen, and her private evidence (the black ball) is never made available to subsequent guessers. If the fourth participant draws a black ball, therefore, that participant also guesses the majority white urn, although the total private evidence is neutral as to the correct urn. Moreover, so does every agent to follow. At least in this example, a cascade is more likely to result in the group guessing correctly than incorrectly, but some will be so-called "reverse" cascades in which the group guesses incorrectly. After all, the amount of evidence that the group actually uses is quite small. When the first two participants draw opposite colored balls, the cascade may be delayed, but it will still occur once there is an imbalance of evidence in favor of one particular urn.

Anderson and Holt ([4]) show that laboratory participants largely behave in the way described above; Hung and Plott ([36]) replicate and extend the results. Easley and Kleinberg ([20]) provide analysis showing that this cascade behavior is indeed rational from a Bayesian perspective, while Baltag et al. ([6]) and Achimescu ([1]) use logical frameworks to explicitly represent the epistemic assumptions of the standard model, confirming the standard analysis of cascades as the result of rational individual choices. Importantly, this holds even under the assumption of common knowledge of rationality. At bottom, since the early (pre-cascade) actions serve as public evidence regarding the color of the urn, and each agent has the same amount of initial private evidence, agents who act on their total evidence are led to act as the previous agents did, and hence to cascade. Rationality is typically taken to require acting on one's total evidence, updating one's beliefs with each new item of evidence, and so cascading is taken to be rational. Note that the agent who updates in this way maximizes the expected accuracy of their beliefs (epistemic utility) and thereby also can better maximize expected utility of their actions.

On the other hand, from a descriptive perspective, we observe that information cascades and crowd-following behavior—though common in real life and the laboratory—are actually much less common and less robust than one would expect from this standard story. Anderson and Holt ([4], p. 852) report that 26% of guesses matched private evidence rather than the total-evidence option when these conflicted —which sometimes forestalled a cascade—in comparison to just 4% of guesses which matched neither. Similarly, Hung and Plott ([36], p. 1517) report that cascades occur only around 78% of the time under the standard set-up, compared to around 97% when conformity rather than truth is rewarded. Perhaps most dramatic is a comparative study of cascade behavior between professional traders (for whom cascades presumably have increased significance) and students. Market professionals were significantly less Bayesian-rational than the students ([2]). Facing a choice between starting a cascade and trusting their private signal in the standard set-up, traders failed to cascade 51% of the time (compared to 39% for students). Despite this, earnings were similar between the two groups, with the traders avoiding more reverse cascades. Furthermore, increased trading intensity among the professionals correlated with less cascading, and the traders as a group learned over the rounds to better avoid reverse cascades. These findings as a whole fit poorly with the idea that rationality requires cascading. Other studies, covering a broader range of potential cascade scenarios, show that people are more willing to trust a cascade the longer it has been going ([40]) and that with longer trials, reverse cascades tend to self-correct ([25, 58]); both findings run contrary to the received view of rational behavior, but result in improved overall performance.[5]

A bit of reflection on everyday life reveals that when cascades occur, not all agents take part. Many people persist in "independent thinking" and non-conformist behavior, for example trying the relatively empty restaurant recommended by a friend, rejecting the novel at the top of the best-seller list in favor of a neglected one with

[5]QRE and cognitive hierarchy models fit the empirical data better than Bayesian models ([25]), but it is not clear that this helps (see e.g. [27]).

an intriguing description, leaving the umbrella at home when everyone outside is carrying one but one's preferred weather website indicates no rain, or even trying to convince others of some fringe hypothesis. Tellingly, many pride themselves on behaving this way.

So there are two components to the standard story. One is unobjectionable from an explanatory standpoint: individuals often follow the crowd, and it is rational for them to do this, even though the group would be more reliable if they instead shared conflicting private information. It is the other component of the story which is problematic: a significant proportion of people (and intuitively many more in real life than in a laboratory setting) behave in accordance with the group good, to their own detriment. Increased experience in relevant situations, and presumably more training in economics, corresponds to less Bayesian rationality. This assessment of the "independent thinkers" raises red flags, as already explained in Section 2.

3.2 Information cascades as Prisoner's Dilemmas

As explained above, the standard analysis of information cascades is an interesting starting point, but not a satisfying final explanation of the empirical data. We must ask the question, why might rational individuals not participate in information cascades? Once we raise the question, a plethora of attractive hypotheses present themselves. In fact, these hypotheses do not need to be invented from scratch; rather, we will show that there is a close analogy between the information cascade problem and the Prisoner's Dilemma.[6] The mountain of literature on the Prisoner's Dilemma therefore provides a list of strategies with proven potential to rationalize apparently-irrational behavior in situations like information cascade problems. We elaborate on the analogy next.

The key to the analogy between information cascades and Prisoner's Dilemmas is that non-cascading individuals are naturally construed as cooperative. They pay a personal cost in reliability because they do not integrate the evidence provided by the earlier actors. Others reap a benefit, namely the additional piece of shared private evidence, which makes them more reliable. At the group level, everyone is better off when the individuals cooperate than when they do not.

		Player 2	
		Cooperate	*Defect*
Player 1	*Cooperate*	$2, 2$	$0, 3$
	Defect	$3, 0$	$1, 1$

Table 1: The Prisoner's Dilemma

[6] Zollman ([73]) discusses and advocates using game theory for social epistemology, as we do here. He gives another example of an epistemic Prisoner's Dilemma.

The Prisoner's Dilemma, depicted in Table 1, is the classic game used for studying such scenarios. In fact, David Coady has noticed the analogy. He remarks in passing that,

> the epistemic virtue of what might be called independent thinking in an information cascade can be compared to the moral virtue of cooperation in prisoner's dilemmas. In both cases everyone involved would be better off if everyone were to behave in a way that is individually irrational ...
> ([16], p. 77)

No-one seems to have followed up on Coady's comment with formal analysis, or checked whether the similarity between the two problems is more than superficial. Our argument in Section 2 shows how valuable this could be, however. As is well known, cooperation is irrational when the Prisoner's Dilemma is played once between two agents acting independently; defecting is rational because from each agent's perspective, this ensures a higher payoff no matter what the other chooses. This does not mean that cooperation is in general irrational, however. It depends on the details of the game, for example whether it will be repeated, if so, how often, and with whom it is played. Precisely because the Prisoner's Dilemma represents a situation in which many people do cooperate and groups are consequently better off—just as in information cascades—the rationality of cooperation has been explored from numerous angles. In this paper, we show how to begin translating this research into the epistemic domain for information cascades.

We begin by spelling out how the information cascade problem can be viewed as a Prisoner's Dilemma in a stricter, formal sense. For concreteness, we continue working with the standard urn scenario. We first define the cooperative strategy, specifying the costs paid and the benefits received. The currency in this setting is epistemic reliability, i.e., likelihood of guessing the correct urn. While any agent could choose to report their private evidence, benefiting those to follow, it is not equally sensible for all agents to do so. For convenience, we name the agents in the information cascade problem according to their order of action in the sequence. We identify Three as the salient agent for forestalling cascades, because this agent is well positioned to improve the reliability of the later agents. Table 2 shows how each agent in a ten-agent sequence is affected by Three's choice to cooperate (report their private evidence) or defect (act on their total evidence).[7] All other agents act on their total evidence.[8] We

[7] For convenience, we simulated the urn scenario to generate the reliability scores; essentially the computer just calculated the probabilities. Private evidence broke ties. Scores are the percentage of correct guesses over 400,000 independent rounds.

[8] We see that agents 5–10 each reap a modest benefit as a result of the third agent's cooperation. That agent 4 *cannot* help later actors by reporting her own private evidence, and would only harm herself, is somewhat surprising but confirmed by additional simulations and pencil-and-paper calculations. For later agents, cooperation makes less sense because there are fewer left to reap the benefits.

therefore define the cooperative strategy as reporting private evidence when in the third position of the sequence, and following the total evidence otherwise.

Expected Accuracy	*One*	*Two*	*Three*	*Four*	*Five*	*Six*	*Seven*	*Eight*	*Nine*	*Ten*
Three Cooperates	67%	67%	67%	74%	79%	79%	81%	81%	82%	82%
Three Defects	67%	67%	74%	74%	77%	77%	79%	79%	79%	79%

Table 2: Reliability Scores

Next, we reduce the payoffs reported in Table 2 into a payoff table with just two symbolic actors, which we could think of as "self" and "other". We first transform the utility functions so that utility of 0 in our analysis represents the baseline (defecting/cascading) reliability, and each percentage point of change in reliability is 1 utile. Then we see from Table 2 that an agent who cooperates as Three gets -7, while the exact benefit to the beneficiary varies slightly based on the beneficiary's position. We idealize away from these marginal differences, assigning the representative beneficiary of cooperation a payoff of +3, so that we can illustrate the incentive structure of our game very simply. Note that agents One, Two, and Four are ignored, since they do not affect and are not affected by any others; from a strategic perspective, they essentially sit out the round.[9]

For the situation to have a Prisoner's Dilemma character, as Coady suggests, requires that any potential cooperator can also expect to be the beneficiary of others' cooperation. Just as in the standard Prisoner's Dilemma, then, rational cooperation requires repetition. The agent's expected payoffs must be a product of not only the costs and benefits, but also the probability with which they pay the costs and reap the benefits. Let p be the probability that any given agent is in the third position, hence in the position of paying the cost, on any given round. Then their expected cost to cooperation is $-7p$, while their expected benefit conditional on others' cooperation is $3(1-p)$. Since information cascades are a group phenomenon, we can expect p to often, or even usually, be small enough that this balance will be favorable. It is also reasonable to assume that these scenarios are often repeated with stable groups; this is true in laboratory settings, and also seems to hold for real-life communities, whether virtual or flesh-and-blood. For example, many people regularly use the same websites and most people move infrequently, so that communities exhibit significant continuity over time.

Although information cascades are a group phenomenon, we can compactly represent the cascade problem using the familiar two-person matrix. This is possible because only one agent (namely Three) has a meaningful choice on any given round, and

[9]For Four, this is so because of evidential ties.

each agent's relative payoff depends only on what Three does. Each agent's expected payoff for the cascade iteration, then, is the product of three factors: the probability that they or some other is agent Three; whether they cooperate as agent Three; and when they are not agent Three, whether the other who is agent Three cooperates. Therefore, from each agent's perspective, they can consider playing the cascade game against a generic opponent, the one whose choice may affect them. This abstraction costs in detail but makes possible a straightforward representation of the game, as shown in Table 3, and renders the problem analytically tractable. We only show Player 1's payoffs for simplicity; the game is symmetric.

		Player 2	
		Cooperate	*Defect*
Player 1	*Cooperate*	$-7p + 3(1-p)$	$-7p$
	Defect	$3(1-p)$	0

Table 3: Information Cascade Game

Setting the familiar backstory aside, what it means for a game to be a Prisoner's Dilemma is just for the players to have particular preferences over the outcomes. Specifically, players must prefer mutual cooperation to mutual defection, but receive their best payoff for defecting when the other cooperates, and their worst payoff in the opposite scenario. The game in Table 3 satisfies these conditions so long as $p < \frac{3}{10}$. Intuitively, then, the repeated cascade scenario is a Prisoner's Dilemma as long as some agent or subgroup does not disproportionately find themselves in the early position. If they do, the other players' actions do not affect them enough, and so they do not act strategically. In the standard laboratory urn set-up, when ten agents all have an equal chance of being in any position in any given round, $p = \frac{1}{7} < \frac{3}{10}$; each agent can expect to be agent Three with an opportunity to cooperate one time for each six times being one of Five, Six, Seven, Eight, Nine, or Ten, and the potential beneficiary of cooperation. Ten agents is a small group as far as information cascades go, but we use this p for illustrative purposes, and reduce the game as depicted in Table 4. This game shows the expected payoffs for each strategy profile when $p = \frac{1}{7}$, now with Player 1's payoffs followed by Player 2's.

		Player 2	
		Cooperate	*Defect*
Player 1	*Cooperate*	$1.57, 1.57$	$-1, 2.57$
	Defect	$2.57, -1$	$0, 0$

Table 4: A Reduced Information Cascade Game

3.3 Rationalizing epistemic cooperation

Section 3.1 considered a concrete information cascade problem and a specific interpretation of that problem as a strategic choice situation. This showed how to create a simple model of the problem in which it is literally a Prisoner's Dilemma. This way of modeling the problem is amenable to rationalizing epistemic cooperation as the result of repeated game strategies, as we discuss presently. In rendering the problem so concrete and, in particular, analytically tractable, many simplifying assumptions were required, however. These greatly limit the real-world conclusions that can be drawn on the basis of that theoretical model alone. Furthermore—as noted earlier—the project of showing how cooperation could be rational in Prisoner's Dilemma-type situations has encompassed many different types of models. Similarly, we think that our simple model is just one way of gaining insight into cooperation in information cascades,[10] and that each modeling strategy will have its own advantages and best range of application. Consequently, we now step back to consider a broader range of strategies for rationalizing cooperation, and discuss how each might apply to information cascades. Since pursuing any one of these strategies would be a substantial project in its own right, our discussion here should be viewed as programmatic, aimed at supporting our general methodological argument and its application to the information cascade problem.

We will discuss kin selection, direct reciprocity, indirect reciprocity, network reciprocity, and multi-level selection as possible mechanisms of rational cooperation, using the framing from Nowak ([50]). Even though evolution is typically conceived of as a competitive process in which the fittest survive, much research shows that evolution can produce cooperation. The key to this is that evolution can align the individuals' competitive interests with the interests of others or the group. Some initial comments are in order. First, Nowak's organization is useful, and much of the study of cooperative behavior has taken an evolutionary perspective, but it is important to note that our interest here is not limited to the question of how epistemic cooperation could have evolved. We are also interested in its rationality from a present-day strategic perspective. Most of the mechanisms we discuss can be understood from either perspective, either as factors influencing the evolution of human traits or as factors influencing the current strategic rationality of employing a cooperative or non-cooperative strategy.

This means, furthermore, that the mechanisms are best understood as ultimate, rather than proximate, explanations ([47]; see also [44, 62] for discussion of the application to the evolution of cooperation literature). Under the evolutionary framing, this means that the mechanisms explain how it could be fitness maximizing to cooperate. Again, though, we can also interpret them non-evolutionarily, and then we would view them as helping to explain how individual cooperation could be utility maximiz-

[10] See [57] for an alternative modeling strategy.

ing (in some salient sense). The proximate explanation may look very different; we comment on other possibilities later, though some discussion is mixed in.

Next, an important caveat pertains especially to the evolutionary interpretation of the mechanisms of cooperation. As Scott-Phillips et al. ([62], pp. 41–42) write, "in particular economic games, cooperative individuals may receive greater payoffs than uncooperative individuals, but these payoffs may not necessarily translate into inclusive fitness effects. Yet this proviso is rarely made explicit." We wish to make it explicit here, as it requires a bit more consideration in the epistemic case as compared to the practical case. The working hypothesis in this literature is that evolution is driven primarily by fitness-based natural selection. The approach that we describe here uses *epistemic reliability* as a proxy for fitness; in the information cascade model in Section 3.2, the probability with which the individual correctly guesses the majority color in the urn is used. Epistemic reliability obviously does not translate perfectly into expected reproductive output, just as the immediate costs and benefits of practical cooperative behaviors do not. We think that it is nonetheless a reasonable proxy (and not less reasonable than the practical analogues). In general, the purpose of epistemic states is to facilitate behavior, and the more accurate the individual's epistemic states, the more appropriate their behaviors will be.[11] In the grand scheme of things, we expect the agents with better epistemic states to also be more reproductively successful. The chain from proxy measure to fitness is a bit longer for epistemic cooperation than for practical cooperation, though, and we think that it is important to remember that this chain is part of any explanation of epistemic cooperation that we generate.

Finally, the mechanisms we discuss here are not mutually exclusive. They should be viewed as available ingredients of explanations of cooperation. Models often combine multiple ingredients, and the most sophisticated studies consider them all.[12] Here, we aim to assess the scope for each ingredient in explaining epistemic cooperation specifically.

The first mechanism of cooperation is kin selection, which we consider only briefly for completeness. The idea is that since genes are the typical unit of natural selection, whether a trait proliferates may be determined only by looking at individual carriers' *inclusive fitness*, which includes both the individual's direct fitness from their own reproductive success and their (weighted) indirect fitness from the reproductive success of relatives ([28]). Thus, it may be (inclusive) fitness maximizing for individuals to cooperate in order to help genetic relatives, even at personal expense. Nonetheless, we need to explain non-cascading behavior in laboratories, websites, and so forth in which the individuals may be perfect strangers. For this purpose, a more general interpretation of kin selection as like-helping-like could play a role.

[11] Modulo Error Management Theory, according to which the relative practical costs of different kinds of false beliefs influence what the agent believes ([29, 30]).

[12] Bowles and Gintis in [12] bring together all of the ingredients as well as a wealth of anthropological data.

The individuals would help others with the same relevant genes, without necessarily being related in the standard sense. This possibility is discussed below.

The second mechanism of cooperation is direct reciprocity, i.e. repeated interactions with the same individuals ([68]). Consider again the Prisoner's Dilemma. While cooperation is generally agreed to be irrational in the classical one-shot version of the game, it is well known that there are cooperative equilibria in the indefinitely repeated version. Cooperative strategies can also achieve the best expected payoffs. As Axelrod and Hamilton ([5]) explain, the most robustly successful strategies are nice (cooperative at the start), retaliatory (punishing defection in kind), and forgiving (resuming cooperation after a punishment round). "Tit for Tat", the best-known strategy for the repeated Prisoner's Dilemma, is especially successful because it embodies these features ([5]). A Tit for Tat player cooperates at the start, and then imitates the opponent's previous choice. See [63] for further discussion of repeated-game strategies.

In our model of an information cascade problem as a Prisoner's Dilemma (Tab. 3), then, strategies (such as Tit for Tat) which result in cooperation every round when played against themselves can be played as best-responses to opponent strategies, resulting in higher payoffs than if the players chose non-cooperative strategies. For this to be the case, continuation of the game needs to be probable enough so that the benefit of current defection is outweighed by the consequently-forgone benefits of future cooperation. Applying the analysis of Roth and Murnighan ([60]) to our model, we find that the continuation probability must be at least .4 for there to be cooperative equilibria, and for Tit for Tat to be an equilibrium strategy when played against itself. Note that the available set of strategies matters to the analysis; Roth and Murnighan consider the strategies of always cooperating, always defecting, Tit for Tat, and "Grim Trigger", which starts by cooperating but defects forever once the opponent does.

Repetition has limitations as a rationalization for cooperation. Apart from the obvious fact that it only applies when fairly stable groups do interact repeatedly, many features of real-world problems can make cooperative repeated-game equilibria less plausible as explanations. For one, implementing a strategy such as Tit for Tat requires that players can tell when another has cooperated or defected. In an information cascade, it will typically be clear when an agent cooperates by going against the crowd, but when they go along with the crowd, it will not always be possible to tell whether they defected or instead their private evidence simply matched that of earlier group members. One can, however, make inferences about how likely an agent is to have cooperated or defected, especially when the group learns the right action at the end and knows the priors behind the private evidence. Another complication is that real agents may make errors in implementing their strategy; cooperation could be destroyed if an accidental defection causes the whole group to defect, so error-responsive strategies must account for the probability of accidental defection. It is harder to see the relevance of this complication for information cascades, since cooperation is (at least in our simple set-up) a much easier strategy than defection. Cooperation just requires the agent to act on their own evidence—perhaps a single piece such as an urn draw—whereas the evidence-integrating calculation to determine the defector strategy could be harder. Finally, cooperative repeated-game equilibria tend to become more rare as the group size grows, for various reasons (see the discussion in [12, Ch.4.5] and the "rules" for direct and indirect reciprocity in [50]). Withdrawing cooperation in response to a defector may in effect punish the entire group, and it may furthermore be hard to tell whether the defection was justified (in response to the other's defection) or unjustified. One can see how this problem would arise in our simple information cascade-Prisoner's Dilemma game from Section 3.2: supposing the third agent went along with the crowd on the last round, should the current third agent cooperate on this round?

The upshot is that in a model including more of these factors, whether cooperative strategies are rational would depend very much on multiple parameter choices, on the chosen strategy space, and on further details. Models of practical cooperation suggest that there is limited scope for repetition as the rationale for actual human cooperation. On the one hand, there are important differences between epistemic cooperation in information cascades and practical cooperation, for example helping to fight in a battle or sharing food. For example, the big-picture costs to voicing an opinion based on less evidence are intuitively lower than the costs of practical cooperation, in general, and in relation to the benefits of an epistemically reliable group. The structure of interaction is also very different, with individuals acting sequentially and taking turns being the cooperator, who benefits multiple group members at once. Detailed models of repeated cascade games will therefore support somewhat different results from their practical analogues. On the other hand, though, we should still expect repetition to explain cooperation only for information cascade situations in which agents' strategies can be observed and repeated game strategies can be implemented reasonably well. The repeated urn problems described in Section 3.1 might be such a case, since the repeated nature of the interaction is clear, behavior is made public, groups are not too large, and when the group is told the correct urn color at the end, it will be possible to draw decent inferences regarding whether the early agents are cooperating or not. Although the group size is much larger and the structure of the problem is a bit rougher, we can also imagine repeated interaction playing an important role in non-cascading among stock market traders. These traders face the same potential cascade problem on every working day, and they have a significant interest in accurately assessing the values of stocks, which is presumably better achieved when much of the group's evidence is integrated as opposed to the opinions of some minority. It would therefore make sense for traders in general to try to avoid jumping on the bandwagon and giving in to fears of crashes or of missing out on an improving stock, if they know that too much cascading behavior would be copied by others to everyone's long-term detriment.

A different mechanism for sustaining cooperation is indirect reciprocity, which is closely connected with reputation. In the preceding discussion, we consider settings in which the same individuals interact repeatedly. In some settings, however, agents repeatedly decide whether to cooperate or defect, but may have no direct experience with the others involved. Instead, they base their choices on the others' track record or reputation as a cooperator or defector. We must note again that a critical difference between standard Prisoner's Dilemmae and information cascades is that cascades are fundamentally a group phenomenon, but different from n-person Prisoner's Dilemmae in that only one individual or a subset are called on to cooperate each round. The line between direct and indirect reciprocity is therefore a bit blurred. We can nonetheless consider the role that reputation might play in epistemic cooperation. Intuitively, reputation plays a key role in individuals' choices to follow the crowd or not. People are often conscious of whether they are seen as independent thinkers, and may work to cultivate this image. This mechanism may be at work in large online communities such as those reviewing books and other products on Amazon.com or similar websites, reviewing restaurants on dining websites, and so forth. The particular group of individuals involved in a cascade problem will differ for each product in question. An agent interacts with one group of people when determining which of the available new Chinese cookbooks to buy, and with a different group when deciding on a performance rain jacket. Although the individuals likely have no prior experience with others in the group, website features may enable everyone to see which other reviewers tend to give especially helpful reviews (note that this removes the cognitive burden and practical challenge of keeping track of individuals' reputations, which is often cited as an obstacle to indirect reciprocity in the practical case). The purpose of the good reputation seems to be different here as compared to the classical models of indirect reciprocity. There, cooperation is meant to secure others' cooperation in the future. On a large website in which many people buy cookbooks and review products, there is no clear way for one agent's current independent purchase and review to prompt others to do the same specifically for products that agent considers buying in the future—especially if we take cooperation to require strictly buying and reviewing something which may be inferior on one's total evidence. The agents could hope to create or support a social norm in which agents in general isolate their decision processes in part from prior ratings and reviews. It is also noteworthy, however, that some websites directly reward very active and helpful reviewers, whether with mere praise and prominence, or with cashback, discounts, or vouchers (as e.g. the major Chinese shopping website Taobao.com does). The benefits of the reputation for independent thinking therefore take a different form from in typical indirect reciprocity models. Proximate causes of cooperation (such as intrinsic desire for such a reputation) also seem to play a critical role here.

A reputation as an epistemic cooperator may also be a valuable ingredient in the next mechanism of cooperation, which Nowak refers to as network reciprocity. For our purposes, the key idea is that in a large group (e.g. on a large website), there may be assortative matching such that cooperative individuals interact more often with other cooperators. If the benefits of mutual cooperation outweigh its costs to any individual, it then pays to be a cooperative "type" to secure access to cooperative rather than non-cooperative groups. From an evolutionary perspective, the types may be hard-wired and the assortative matching may arise because offspring inherit their type from their parents, stay close by, and individuals interact primarily with those around them. In current epistemic communities, we more often see that individuals choose with whom to interact, for example on the basis of reputation. Within internet forums and social media communities, for example, individuals can decide who to follow, which posts to comment on and to share, and which individuals to ignore, chastise, censor, or even ban from their groups. A group of epistemic cooperators can take steps to support cooperators and reduce cascade-type behavior within their group. For example, new or different views might be engaged with or up-voted by others who agree with it, while those who parrot what others have said can have their comments removed as "duplicates" on some platforms. In contrast, a more conformist group may essentially do the opposite, supporting repetitive posts and ignoring, down-voting, etc. unwelcome new views.

A more thoroughly evolutionary mechanism for generating and sustaining cooperation is group selection, or perhaps better named, multi-level selection. Bowles and Gintis ([12]) compile evidence that violent inter-group conflict was frequent in human evolutionary history, and that it played an important role in the selection process leading to modern humans. They construct models in which cooperation can spread because it makes groups more likely to survive and spread, for example by defeating less cooperative groups in war, even if cooperative individuals are worse off within each group. It is easy to see how epistemic cooperation would be beneficial to groups, too; a wider pool of evidence, more possibilities investigated, or a broader range of suggestions for consideration could help the group to reach better judgments regarding critical issues such as food, shelter and engagement with other groups. Groups who avoid reverse information cascades (e.g. thanks to cooperators) are more epistemically reliable, and we expect epistemic reliability to be practically beneficial to groups just as to individuals. As Okasha explains, well-functioning higher-level agents (such as groups, here) tend to have ways to suppress competition and conflict between their parts (here, individual people, see [53], Ch. 1.6). A group could best promote its epistemic interests by creating institutions counter-acting the individual costs of cooperation, for example. Bowles and Gintis discuss "reproductive leveling" institutions such as food sharing which reduce the fitness differences between cooperators and non-cooperators, thus supporting the persistence of cooperators in the population. Such practices would presumably have the same impact on epistemic forms of cooperation. One can also imagine groups having rules or practices specifically to reduce the individual costs to independent thinking, for example having multiple rounds of opinion or evidence sharing before action must be taken, making important decisions together as a group, rewarding dissenters, etc.

In terms of information cascades, such fundamentally evolutionary explanations of cooperation seem to best support general tendencies to cooperate or to defect. Insofar as selection (whether for epistemically reliable groups or free-thinking individuals) is responsible for independent thinking, the tendency to do so would probably extend beyond information cascade situations specifically and manifest as a general way of responding to social evidence. In fact, we perceive the disposition to think independently to indeed be very general and widespread, and find it to be exceedingly plausible that this has a general evolutionary rationale. It is also possible, however, to imagine group-level pressures playing a role in the present maintenance of epistemic cooperation. In many cascade scenarios, such as laboratory urn trials or product purchasing on large websites, the group does not necessarily conceive of itself as a unit with an interest in its reliability. In other cases, however, the group exhibits a "unity of purpose" ([53], Ch. 1.6), and if the group realizes the value of epistemic cooperation, it may take steps to promote it, for example in the ways suggested above. Well-run companies and other organizations may do this, for example. The present-day analog of group selection would then be that organizations which manage to cultivate epistemic cooperation tend to survive, expand, and reproduce themselves to a greater extent than do organizations which are beset by information cascades and similar problems.

As we noted earlier, the five mechanisms supporting cooperation that we have discussed are primarily meant to provide ultimate explanations of cooperation. However, as Marchionni and Vromen ([44]) discuss (and as can be seen in the preceding discussion), proximate causes are sometimes an important part of the story; even an ultimate explanation may depend on the presence of proximate mechanisms. We close this section by elaborating on the proximate causes or enablers of cooperation, some of which have been mentioned above. Firstly, human culture has developed along with us, and many aspects of our culture can support independent thinking.[13] Institutions are known to be important to sustaining practical cooperation (see [54] for an overview of the research on collective action problems). For example, we often reward or praise those who express minority opinions which turn out to be true. Secondly, specific capacities underly our abitily to sustain cooperation. In the practical domain, these include the (logical) ability to detect cheaters ([17]), the ability to keep track of others' past behavior and reputations, and the ability of cooperators to recognize other cooperators ([13]) or to recognize "similar" agents with whom to cooperate ([59]). In addition to the more general abilities, we clearly seem to have

[13] Scott-Phillips et. al. ([62]) argue that cultural evolution should be seen as a proximate mechanism, though this is not uncontroversial. Bowles and Gintis ([12]) and others build models in which nature (genes or phenotypes) and culture co-evolve.

ways of identifying epistemic cooperators in our communities; they often make themselves known, for example by choice of language, case and punctuation on the internet. Finally, various authors have emphasized that evolutionarily-indicated strategies may be implemented by the mind via emotions, internalized social and moral norms, and other simple heuristics ([8, 31, 35, 39]). For example, people apparently enjoy punishing those who violate social norms for its own sake ([55]). These proximate mechanisms would be no less relevant to epistemic cooperation than to practical cooperation, and the story also matches our experience. Many people would be ashamed to think of themselves as crowd followers, and we seem to regularly apply simple heuristics such as to treat with suspicion or discount opinions of others that do not match our personal evidence or experience.

It should be clear from the foregoing discussion that to characterize non-cascading individuals in information cascades as irrational is much too quick. These individuals may be less epistemically reliable from a short-term perspective, but we will not be able to explain their behavior, or to determine whether it is rational in interesting senses, until we explore the various rationales for cooperation as has been done in the practical case. While this section has focused on information cascades as a high-profile phenomenon and a clear-cut case of the methodological problem discussed in Section 2, we think that epistemic cooperation extends far beyond problems with the formal cascade structure, and that the explanatory hypotheses discussed in this section have broader applicability.

4 Conclusion

We end by briefly reviewing our arguments and main conclusions here. First, although the two parts of the theoretical Independence Thesis are symmetric, there remains an important asymmetry from an explanatory perspective. We should not be surprised when we find real-world cases in which individuals behave rationally, but the individuals' behaviors fail to coordinate for the good of the group. In contrast, we should be suspicious of analyses of real-world cases according to which individuals are irrational but groups are successful, especially insofar as the individual irrationality is widely and systematically responsible for the group success. The literature on rationalizing practical cooperation is a shining example of this lesson being taken seriously, thus greatly enriching our understanding.

Given our broad conception of rationality, this position is in line with both methodological literature and social science practice. Nonetheless, the problematic character of certain explanations has been neglected in a truly major areas of research—information cascades (in which the standard analysis can only support the problematic type of explanation). The fact that the explanandum in this case is group epistemic reliability rather than practical success does not make a difference in terms of the standards for good explanations.

We argue, however, that the solution is at hand: it is enlightening to see these seemingly-distinct research problems as instances of a common problem, namely that of rationalizing epistemic cooperation on the part of individuals. Given this perspective, the paradigmatic example of the project of rationalizing practical cooperation (as in the Prisoner's Dilemma literature) is shown to be valuable in a more concrete sense, as a model for how to rationalize cooperation in the case we discuss. This familiar literature provides a useful overarching perspective on these epistemic choice problems, a list of strategies for analyzing and understanding them, and caveats such as the distinction between ultimate and proximate explanations. Moving forward, we hope to see more emphasis on the potential rationality of epistemic cooperators, more recognition of its critical social value wherever it appears (even when it may take forms we do not like, as when we find minority opinions unpalatable), greater appreciation for and attention to the apparently deep-seated nature of human epistemic non-conformity, and ultimately efforts to channel it for the improved social outcomes. Without proper understanding of independent thinking, the latter would be considerably less likely to succeed.

References

[1] A. C. Achimescu, 2014, *Games and Logics for Informational Cascades*, mathesis, University of Amsterdam.

[2] J. E. Alevy, M. S. Haigh and J. A. List, 2007, "Information cascades: Evidence from a field experiment with financial market professionals", *The Journal of Finance*, **62(1)**: 151–180.

[3] M. Allais, 1953, "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine", *Econometrica*, **21(4)**: 503–546.

[4] L. Anderson and C. Holt, 1997, "Information cascades in the laboratory", *The American Economic Review*, **87(5)**: 847–862.

[5] R. Axelrod and W. D. Hamilton, 1981, "The evolution of cooperation", *Science*, **211(4489)**: 1390–1396.

[6] A. Baltag, Z. Christoff, J. Hansen and S. Smets, 2013, "Logical models of informational cascades", in J. van Benthem and F. Liu (eds.), *Logic Across the University: Foundations and Applications*, pp. 405–432, London: College Publications.

[7] A. V. Banerjee, 1992, "A simple model of herd behavior", *The Quarterly Journal of Economics*, **107(3)**: 797–817.

[8] W. M. Bennis, D. L. Medin and D. M. Bartels, 2010, "The costs and benefits of calculation and moral rules", *Perspectives on Psychological Science*, **5(2)**: 187–202.

[9] N. Berg, 2014, "The consistency and ecological rationality approaches to normative bounded rationality", *Journal of Economic Methodology*, **21(4)**: 375–395.

[10] S. Bikhchandani, D. Hirshleifer and I. Welch, 1992, "A theory of fads, fashion, custom, and cultural change as informational cascades", *The Journal of Political Economy*, **100(5)**: 992–1026.

[11] J. Birch, 2014, "How cooperation became the norm (essay review of *Cooperation and Its Evolution*, ed. K. Sterelny et al.)" *Biology and Philosophy*, **29(3)**: 433–444.

[12] S. Bowles and H. Gintis, 2013, *A Cooperative Species: Human Reciprocity and Its evolution*, Princeton: Princeton University Press.

[13] J. Brosig, 2002, "Identifying cooperative behavior: Some experimental results in a prisoner's dilemma game", *Journal of Economic Behavior & Organization*, **47(3)**: 275–290.

[14] L. Buchak, 2013, *Risk and Rationality*, Oxford: Oxford University Press.

[15] N. Cartwright, 1983, *How the Laws of Physics Lie*, Oxford: Claredon Press.

[16] D. Coady, 2006, "When experts disagree", *Episteme*, **3(2)**: 68–79.

[17] L. Cosmides, 1989, "The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task", *Cognition*, **31(3)**: 187–276.

[18] D. Davidson, 1974, "Psychology as philosophy", in S. Brown (ed.), *Philosophy of Psychology*, pp. 41–52, London: Palgrave Macmillan.

[19] R. M. Dawes, 1980, "Social dilemmas", *Annual Review of Psychology*, **31(1)**: 169–193.

[20] D. Easley and J. Kleinberg, 2010, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, New York: Cambridge University Press.

[21] D. Ellsberg, 1961, "Risk, ambiguity, and the Savage axioms", *Quarterly Journal of Economics*, **75(2)**: 643–669.

[22] L. G. Epstein and J. Zhang, 2001, "Subjective probabilities on subjectively unambiguous events", *Econometrica*, **69(2)**: 265–306.

[23] R. H. Frank, 2011, *The Darwin Economy: Liberty, Competition and the Common Good*, Princeton: Princeton University Press.

[24] G. Gigerenzer, 2000, *Adaptive Thinking: Rationality in the Real World*, Oxford: Oxford University Press.

[25] J. K. Goeree, T. R. Palfrey, B. W. Rogers and R. D. McKelvey, 2007, "Self-correcting information cascades", *The Review of Economic Studies*, **74(3)**: 733–762.

[26] F. Gul and W. Pesendorfer, 2008, "The case for mindless economics", in A. Caplin and A. Schotter (eds.), *The Foundations of Positive and Normative Economics: A Handbook*, pp. 3–42, New York: Oxford University Press.

[27] P. A. Haile, A. Hortaçsu and G. Kosenok, 2008, "On the empirical content of quantal response equilibrium", *American Economic Review*, **98(1)**: 180–200.

[28] W. D. Hamilton, 1964, "The genetical evolution of social behavior, I & II", *Journal of Theoretical Biology*, **7(1)**: 1–52.

[29] M. G. Haselton and D. M. Buss, 2000, "Error management theory: A new perspective on biases in cross-sex mind reading", *Journal of Personality and Social Psychology*, **78(1)**: 81–91.

[30] M. G. Haselton and D. Nettle, 2006, "The paranoid optimist: An integrative evolutionary model of cognitive biases", *Personality and Social Psychology Review*, **10(1)**: 47–66.

[31] R. A. Heiner, 1990, "Rule-governed behavior in evolution and human society", *Constitutional Political Economy*, **1(1)**: 19–46.

[32] V. F. Hendricks and P. G. Hansen, 2016, *Infostorms: Why Do We 'Like'? Explaining Individual Behavior on the Social Net*, Cham: Springer.

[33] D. Hershleifer and S. H. Teoh, 2003, "Herd behavior and cascading in capital markets", *European Financial Management*, **9(1)**: 25–66.

[34] R. Hertwig and G. Gigerenzer, 1999, "The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors", *Journal of Behavioral Decision Making*, **12(4)**: 275–305.

[35] R. Hertwig and U. Hoffrage, 2013, "Simple heuristics: The foundations of adaptive social behavior", in R. H. et al. (ed.), *Simple Heuristics in a Social World*, pp. 3–36, New York: Oxford University Press.

[36] A. A. Hung and C. R. Plott, 2001, "Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions", *American Economic Review*, **91(5)**: 1508–1520.

[37] J. B. Kadane, 1992, "Healthy scepticism as an expected-utility explanation of the phenomena of Allais and Ellsberg", *Theory and Decision*, **32(1)**: 57–64.

[38] K. V. Katsikopoulos, 2014, "Bounded rationality: The two cultures", *Journal of Economic Methodology*, **21(4)**: 361–374.

[39] M. M. Krasnow and A. W. Delton, 2016, "Are humans too generous and too punitive? Using psychological principles to further debates about human social evolution", *Frontiers in Psychology*, **7**, 799, https://doi.org/10.3389/fpsyg.2016.00799.

[40] D. Kübler and G. Weizsäcker, 2005, "Are longer cascades more stable?", *Journal of the European Economic Association*, **3(2-3)**: 330–339.

[41] C. J. Lee, 2010, "Reclaiming Davidson's methodological rationalism as Galilean idealization in psychology", *Philosophy of the Social Sciences*, **40(1)**: 84–106.

[42] C. List and P. Pettit, 2002, "Aggregating sets of judgments: An impossibility result", *Economics and Philosophy*, **18(1)**: 89–110.

[43] S. Lohmann, 2019, "Bandwagon effects, information cascades, and the power in numbers", in E. Suhay, B. Grofman and A. H. Trechsel (eds.), *The Oxford Handbook of Electoral Persuasion*, pp. 730–753, Oxford: Oxford University Press.

[44] C. Marchionni and J. Vromen, 2009, "The ultimate/proximate distinction in recent accounts of human cooperation", *Tijdschrift voor Filosofie*, **71(1)**: 87–117.

[45] C. Mayo-Wilson, K. J. S. Zollman and D. Danks, 2011, "The independence thesis: When individual and social epistemology diverge", *Philosophy of Science*, **78(4)**: 653–677.

[46] C. Mayo-Wilson, K. J. S. Zollman and D. Danks, 2013, "Wisdom of crowds versus groupthink: Learning in groups and in isolation", *International Journal of Game Theory*, **42(3)**: 695–723.

[47] E. Mayr, 1961, "Cause and effect in biology", *Science*, **134(3489)**: 1501–1506.

[48] E. McMullin, 1985, "Galilean idealization", *Studies in History and Philosophy of Science Part A*, **16(3)**: 247–273.

[49] N. I. Al-Najjar and J. Weinstein, 2009, "The ambiguity aversion literature: A critical assessment", *Economics & Philosophy*, **25(3)**: 249–284.

[50] M. A. Nowak, 2006, "Five rules for the evolution of cooperation", *Science*, **314(5805)**: 1560–1563.

[51] M. Oaksford and N. Chater, 1994, "A rational analysis of the selection task as optimal data selection", *Psychological Review*, **101(4)**: 608–631.

[52] S. Okasha, 2006, *Evolution and the Levels of Selection*, Oxford: Oxford University Press.

[53] S. Okasha, 2018, *Agents and Goals in Evolution*, New York: Oxford University Press.

[54] E. Ostrom, 2000, "Collective action and the evolution of social norms", *Journal of economic perspectives*, **14(3)**: 137–158.

[55] D. J.-F. de Quervain et al., 2004, "The neural basis of altruistic punishment", *Science*, **305(5688)**: 1254–1258.

[56] P. Rich, 2016, "Axiomatic and ecological rationality: Choosing costs and benefits", *Erasmus Journal for Philosophy and Economics*, **9(2)**: 90–122.

[57] P. Rich, "Hidden costs of epistemic conformity: Lessons from information cascade simulations", *British Journal for Philosophy of Science*, forthcoming, https://doi.org/10.1086/717292.

[58] A. van de Rijt, 2019, "Self-correcting dynamics in social influence processes", *American Journal of Sociology*, **124(5)**: 1468–1495.

[59] R. L. Riolo, M. D. Cohen and R. Axelrod, 2001, "Evolution of cooperation without reciprocity", *Nature*, **414(6862)**: 441–443.

[60] A. E. Roth and J. K. Murnighan, 1978, "Equilibrium behavior and repeated play of the Prisoner's Dilemma", *Journal of Mathematical Psychology*, **17(2)**: 189–198.

[61] R. Samuels, S. Stich and M. Bishop, 2002, "Ending the rationality wars: How to make disputes about human rationality disappear", in R. Elio (ed.), *Common Sense, Reasoning and Rationality*, pp. 236–268, New York: Oxford University Press.

[62] T. C. Scott-Phillips, T. E. Dickins and S. A. West, 2011, "Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences", *Perspectives on Psychological Science*, **6(1)**: 38–47.

[63] J. Sell and R. K. Wilson, 1999, "The maintenance of cooperation: Expectations of future interaction and the trigger of group punishment", *Social Forces*, **77(4)**: 1551–1571.

[64] E. Stein, 1996, *Without Good Reason: The Rationality Debate in Philosophy and Science*, Oxford: Oxford University Press.

[65] K. Stenning and M. Van Lambalgen, 2012, *Human Reasoning and Cognitive Science*, Cambridge, MA: MIT Press.

[66] S. Stich, 1993, "Naturalizing epistemology: Quine, Simon and the prospects for pragmatism", in C. Hookway and D. Peterson (eds.), *Royal Institute of Philosophy Supplements, **Vol. 34**: Philosophy and Cognitive Science*, pp. 1–17, Cambridge: Cambridge University Press.

[67] T. Sturm, 2012, "The 'rationality wars' in psychology: Where they are and where they could go", *Inquiry*, **55(1)**: 66–81.

[68] R. L. Trivers, 1971, "The evolution of reciprocal altruism", *The Quarterly review of biology*, **46(1)**: 35–57.

[69] A. Tversky and D. Kahneman, 1983, "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment", *Psychological Review*, **90(4)**: 293–315.

[70] A. Wallin, 2013, "A peace treaty for the rationality wars? External validity and its relation to normative and descriptive theories of rationality", *Theory & Psychology*, **23(4)**: 458–478.

[71] P. C. Wason, 1968, "Reasoning about a rule", *Quarterly Journal of Experimental Psychology*, **20(3)**: 273–281.

[72] M. Weber, 1947, *The Theory of Economic and Social Organization*, New York: Oxford University Press.

[73] K. J. Zollman, 2021, "The theory of games as a tool for the social epistemologist", *Philosophical Studies*, **178(4)**: 1381–1401.

实践中的独立性论题：如何理解群体认知成功

帕翠夏·里奇　　刘海林 ✉

摘　要

社会认识论的一个重要经验是由两部分内容构成的“独立性论题”：群体可能是非理性的，尽管其每一成员都是理性的；群体也可能是理性的，尽管其成员都是非理性的。我们强烈建议将合作合理化方面的研究视为一个范例，进而对现实世界中涉及论题中第二部分内容的分析提出质疑。我们集中讨论了这类分析的一个突出例子：信息层叠，并论证说明它可以富有成效地概念重构为认知合作的一个实例。这一新视角为改进对这一重要情景的解释和理解指明了方向。

帕翠夏·里奇　拜罗伊特大学哲学系
patricia.rich@uni-bayreuth.de
刘海林 ✉　中山大学逻辑与认知研究所
中山大学哲学系
liuhlin3@mail.sysu.edu.cn