RMM Vol. 0, Perspectives in Moral Science,
ed. by M. Baurmann & B. Lahno, 2009, 49–65
http://www.rmm-journal.de/

Max Albert

Why Bayesian Rationality Is Empty, Perfect Rationality Doesn't Exist, Ecological Rationality Is Too Simple, and Critical Rationality Does the Job*

Abstract:

Economists claim that principles of rationality are normative principles. Nevertheless, they go on to explain why it is in a person's own interest to be rational. If this were true, being rational itself would be a means to an end, and rationality could be interpreted in a non-normative or naturalistic way. The alternative is not attractive: if the only argument in favor of principles of rationality were their intrinsic appeal, a commitment to rationality would be irrational, making the notion of rationality self-defeating. A comprehensive conception of rationality should recommend itself: it should be rational to be rational. Moreover, since rational action requires rational beliefs concerning means-ends relations, a naturalistic conception of rationality has to cover rational belief formation including the belief that it is rational to be rational. The paper considers four conceptions of rationality and asks whether they can deliver the goods: Bayesianism, perfect rationality (just in case that it differs from Bayesianism), ecological rationality (as a version of bounded rationality), and critical rationality, the conception of rationality characterizing critical rationalism. The answer is summarized in the paper's title.

1. Rationality in Economics

In economics, it is often assumed that rationality is described by a set of normative principles. Consider the transitivity axiom for preferences, which belongs to the basic principles of rationality in economics. Of course, economists know that, actually, people often violate the transitivity axiom. They consider the transitivity axiom as a theoretical idealization and, at the same time, a normative ideal: a person's preferences ought to be transitive. And it seems that most people accept this norm. Thus, experimental subjects who are confronted with the fact

* This paper is based on a talk at the 25th Roland Seminar at the Max Planck Institute of Economics in Jena 2008, where Hartmut Kliemt invited me to speak on "rationality in connection with experimental economics". I have profited from discussions with seminar participants, especially Michael Baurmann, Werner Güth, Susanne Hahn, and Bernd Lahno. My greatest and longstanding debts, however, are to Hartmut Kliemt, with whom I have been discussing economics and philosophy since my student days when I first met him in Alpbach.

that they have stated intransitive preferences typically concede that they have made a mistake and correct themselves.

But economists usually do not leave it at that. They present a further argument in favor of transitivity, explaining why it is in a person's own interest to avoid intransitive preferences. This is the famous money-pump argument, which shows that a person with intransitive preferences can, in the worst case, be exploited systematically through trade.[1]

The money-pump argument, however, suggests that there is nothing normative about the transitivity axiom. According to the argument, transitivity protects against losses and, consequently, is a means for achieving an end. Statements about the relation between means and ends are not normative. It is either true or false that transitivity offers such a protection; this is not a question of norms or value judgments. If people want to protect themselves and believe that transitive preferences are an effective means of protection, it is rational for them to correct any intransitivities in their preferences, at least if no other of their goals are affected.[2] Transitivity as a normative requirement is superfluous.

The money-pump argument suggests that the conception of rationality used in economics might be naturalistic, that is, non-normative. Indeed, economists should hope that a naturalistic interpretation of rationality is possible because the alternative is not very attractive. Assume that the only argument in favor of principles of rationality would be their intrinsic appeal. We would then have to say that a commitment to rationality is irrational, which makes the notion of rationality self-defeating. A comprehensive conception of rationality should recommend itself, that is, it should be rational to be rational (cf. also Bartley 1987; Musgrave 1993, 294–297; 1999, 329–331, 336–337).[3]

There are two objections against a naturalistic conception of rationality. First, it is sometimes argued that there is one normative principle that cannot be avoided in a discussion of rationality, namely, that one should be rational (see Wallace 2009, section 4). Second, it is argued that rational action requires that

[1] See Kliemt 2009, 45, where the money-pump argument (though not under this name) is used to characterize intransitivity as pragmatic incoherence. The possibility of systematic exploitation just adds drama to the argument; trading with several well-meaning partners can lead to the same result.

[2] The tension between the idea of given preferences in economics and arguments like the money-pump argument can be dissolved if preferences are viewed as summarizing all considerations a person considers to be relevant in a decision between different options (Kliemt 2009, section 3.2). According to this concept of "preferences all things considered" (or "satiated preferences", see Kliemt 2009, 124), transitivity is the result of deliberation and can be viewed as a means to an end, as the money-pump argument assumes.

[3] There is an easy way, however, to make such a conception of rationality comprehensive. Consider the rather silly normative principle that it is rational to accept all intrinsically appealing principles. If this principle intrinsically appeals to you, it recommends itself. Let us call this version of rationality 'uncritical rationality'. Uncritical rationality requires that you accept all other principles you find intrinsically appealing, for instance, transitivity of preferences. Uncritical rationality is analogous to the epistemology of rationalism with its notion of self-evident truth. The challenge for a normative conception of rationality is to find a self-recommending version that is better than uncritical rationality.

the beliefs concerning means-ends relations are rational, and that normative principles determine the rationality of beliefs.

The first objection seems to me unfounded. Consider Indiana Jones at the Shanghai night club where he has been poisened but finally manages to get his hands on the phial with the antidot. He has good reasons—well, not really, but let's pretend—for believing that he will survive if and only if he swallows the antidot; moreover, he wants to survive, and there are no other relevant goals or beliefs. This motivates him to swallow the antidot. He needs no argument with the normative conclusion that he ought to do so. Such an argument would indeed require some normative principle as a premise. However, a theory of rational action can do without such an argument because, at some point, it has to assume anyway that people are motivated to act. Goals and beliefs are sufficient as a motivation; normative principles are superfluous.

The same goes for a conception of rationality. When we say that swallowing the antidot is rational for Indiana Jones, this just means that he has good reasons to believe that this is an effective means for reaching his goals.[4]

The second objection against a naturalistic conception of rationality poses more difficulties. It would not be rational for Indiana Jones to swallow the antidot if he had no good reasons for believing that this was an effective means for reaching his goals. We therefore have to extend rationality to beliefs.

Traditionally, the rationality of beliefs is viewed as a problem of theoretical rationality, while the rational choice of actions falls into the realm of practical rationality (Wallace 2009). However, it seems to me that we could view both, belief formation and the choice of actions, as decision problems. A relevant hypothesis can be viewed as a solution proposal to the decision problem of what to believe in a certain situation. Thus, an effective means for distinguishing true from false hypotheses would solve the problem of rational belief formation in a naturalistic way.

It is, of course, clear that there can be no sure-fire method for finding true beliefs. Any method for forming beliefs must be a heuristic in the sense that it cannot deliver true beliefs with certainty or even with some known objective probability.[5] All we can hope for are heuristics with good working properties.

In this paper, I will consider four conceptions of rationality and ask whether they can do the job described above: Bayesianism, perfect rationality (just in case that it differs from Bayesianism), ecological rationality (as a version of bounded rationality), and critical rationality, the conception of rationality characterizing critical rationalism. The focus will be on belief formation, starting from the problem of induction. The results of the paper are already summarized in its title.

---

[4] Thus, as argued by Kliemt, the participant's attitude, according to which we place ourselves in the shoes of others, and the objective attitude of a rational-choice explanation with goals and beliefs as initial conditions are complements; cf. Kliemt 2009, section 2.3, especially 31–32.

[5] This follows from modern epistemology; see Musgrave 1993 for an introduction.

2. Bayesian Rationality

2.1 A Simple Illustration of the Problem of Induction

Irrationalism, or radical skepticism, is a resignative answer to the problem of induction. Irrationalists argue that, since the past teaches us nothing about the future, any beliefs and any decisions are as rational or irrational as any other, no matter what our goals and experiences are.

As a simple illustration of this point, consider the set of all possible worlds, illustrated by the rectangle in figure 1, where the probabilities can be ignored for now. The actual world is a point in the rectangle; each other point is another possible world. Before any experience, we cannot know which of the infinitely many possible worlds is the actual world.

Events are represented by subsets of the set of possible worlds. Each event corresponds to the set of those possible worlds where the event occurs. Let X and Y be two variables denoting observable events. First, we will observe whether $X = 0$ or $X = 1$; later, we will observe whether $Y = 0$ or $Y = 1$. Each variable corresponds to a partitioning of the rectangle; there are four sets of possible worlds, each characterized by one possible realization of (X, Y).

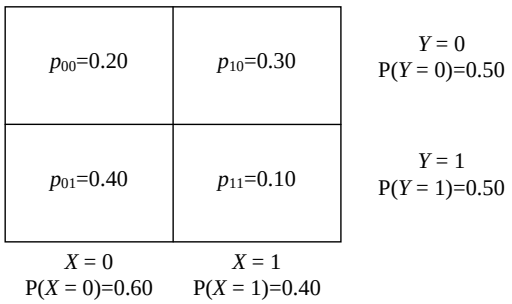

Figure 1: Set of all posible worlds partitioned into four quadrants, corresponding to the four possible realizations of (X, Y), and prior probability distribution P with $P(X = i, Y = j) = p_{ij}, i, j = 0, 1$.

Once we have observed X, we know whether the actual world is on the left-hand side or the right-hand side in figure 1. However, this tells us nothing about Y, that is, whether the actual world is in the top or in the bottom part. Thus, we may believe what we want about Y (the future) after having observed X (the past). Any beliefs about the future, and, consequently, any decisions are as rational or irrational as any other, no matter what our goals and experiences are. This is the problem of induction.

2.2 The Bayesian Recipe

Bayesianism tries to solve the problem of induction with the help of the probability calculus.[6] The following recipe describes the Bayesian account of rational learning and decision making.

1. *Select prior probabilities:* Choose a prior probability distribution on the set of all possible worlds. From this distribution, derive probabilities for future events and use them for predictions and decision making.
2. *Compute posterior probabilities:* After observing some event, adopt a new probability distribution called posterior probability distribution. The posterior probabilities for all future events are equal to the prior probabilities conditional on the observation. Henceforth, use the posterior probabilities for predictions and decision making.[7]
3. *Repeat ad infinitum:* Any further observation leads to further conditioning. The posterior probabilities at one stage serve as prior probabilities at the next. This process of updating probabilites on the basis of observations is called Bayesian learning.

Step 1 of the Bayesian recipe calls for the choice of a prior distribution, subsequently denoted by P. Imagine this as a heap of whipped cream distributed on the surface of the rectangle in figure 1. The distribution may be very uneven, with only a thin spread or no cream at all at one place and big heaps at other places. The probability of each event is the share of cream covering the corresponding set of possible worlds. Let us assume that the shares of cream on the four quadrants are as given by the numbers p_{ij} in figure 1. Thus, 30% of all the cream is on the top-right quarter, meaning that $P(X = 1, Y = 0) = p_{10} = 0.30$.

The initial choice of a prior probability distribution is not regulated in any way. The probabilities, called subjective or personal probabilities, reflect personal degrees of belief. From a Bayesian philosopher's point of view, any prior distribution is as good as any other. Of course, from a Bayesian decision maker's point of view, his own beliefs, as expressed in his prior distribution, may be better than any other beliefs, but Bayesianism provides no means of justifying this position. Bayesian rationality rests in the recipe alone, and the choice of the prior probability distribution is arbitrary as far as the issue of rationality is concerned. Thus, two rational persons with the same goals may adopt prior distributions that are wildly different.

Step 2 of the Bayesian recipe uses the concept of a conditional probability, that is, the probability of one event, say, $Y = 0$, on the condition that some other event, say, $X = 1$, occurs. The conditional probability of $Y = 0$ given $X = 1$ is denoted by $P(Y = 0|X = 1)$. In terms of cream, we ask which percentage of

[6] For expositions and critical discussions of Bayesianism, see Albert 2001, 2003; Binmore 2009; Earman 1992; Gillies 2001; and Howson and Urbach 1993. In this paper, I am only concerned with subjective Bayesianism. Albert (2003), Earman (1992, 139–141) and Howson and Urbach (1993, ch. 4 and 15i) also criticize objective Bayesianism.

[7] If the observed event has a prior probability of zero, conditional probabilities are not defined, and the posterior distribution can be chosen freely. I do not consider this possibility explicitly, as this would only complicate the exposition without changing the conclusions.

the cream covering the right-hand side (representing $X = 1$) is in the upper part (representing $Y = 0$). According to figure 1, 40% of the cream is on the right-hand side and 30% of the cream is in the top-right quarter. Therefore, 30/40 × 100% = 75% of the right-hand side's cream is in the upper part; hence, $P(Y = 0|X = 1) = p_{10}/(p_{10} + p_{11}) = 0.75$.

Bayesian learning requires that, after observing the value of X, we adopt the respective conditional probabilities as posterior probabilities, subsequently denoted by P_{new}. Let us assume that we observe $X = 1$. We know, then, that the actual world must be on the right-hand side. Our new probabilities are the old probabilities conditional on the event we have observed: $P_{new}(Y = 0) = P(Y = 0|X = 1) = 0.75$.

In terms of cream, Bayesian learning implies that cream never moves. Once we have learned that the actual world must be on the right-hand side, the left-hand side and all the cream on it are just forgotten. All that counts for further considerations is the right-hand side with the original amount and distribution of cream.

2.3 Back to Irrationalism

Given the prior probability distribution of figure 1, the Bayesian recipe forces us to conclude that the probability of $Y = 0$ rises from 0.50 to 0.75 if we observe $X = 1$. According to this recipe, then, we are no longer free in our beliefs after observing X. Does this mean that the problem of induction has been solved?

In order to answer this question, we have to distinguish between flexibility and arbitrariness. Bayesian learning is completely inflexible after the initial choice of probabilities: all beliefs that result from new observations have been fixed in advance. This holds because the new probabilities are just equal to certain old conditional probabilities (or, in terms of cream, because cream never moves).

However, the problem of induction is the arbitrariness of beliefs. Inflexibility may be accompanied by arbitrariness or not, depending on whether each and any sequence of beliefs could be fixed in advance or whether at least some sequences are ruled out. According to the Bayesian recipe, the initial choice of a prior probability distribution is arbitrary. But the probability calculus might still rule out some sequences of beliefs and thus prevent complete arbitrariness.

Actually, however, this is not the case: nothing is ruled out by the probability calculus. The example of figure 1 shows why. We have $P(X = 1) = p_{10} + p_{11}$ and $P_{new}(Y = 0) = p_{10}/(p_{10} + p_{11})$. If we are free to choose p_{10} andp_{11} in any way we want, beliefs concerning Y after observing $X = 1$ are completely independent from beliefs concerning X. In terms of cream: the decision of how cream should be distributed between left and right is independent from the decision of how, on each side, it is distributed between top and bottom.

Thus, anything goes. Observing $X = 1$ can have any impact on our probability for $Y = 0$ we like it to have. The past does not tell us anything about the future; hence, probabilities of future events are independent from past events

and their probabilities. Bayesian rationality allows us to learn what we want from the past. By adopting a suitable prior probability distribution, we can fix the consequences of any observations for our beliefs in any way we want. This result, which will be referred to as the anything-goes theorem, holds for arbitrarily complicated cases and any number of observations. It implies, among other consequences, that two rational persons with the same goals and experiences can, in all eternity, differ arbitrarily in their beliefs about future events.[8]

Keeping to the Bayesian recipe, then, cannot, by and in itself, help us make better decisions. It just burdens us with a lot of calculations. Moreover, it seems that there is nothing else we get from it. Indeed, no matter what we do and what we would like to achieve, we can rest assured in the certain knowledge—at least as far as certainty is to be had in mathematics—that there exists a prior probability distribution that rationalizes our beliefs and actions, and even all our contingent plans, in terms of Bayesian learning.

From a Bayesian point of view, any beliefs and, consequently, any decisions are as rational or irrational as any other, no matter what our goals and experiences are. Bayesian rationality is just a probabilistic version of irrationalism. Bayesians might say that somebody is rational only if he actually rationalizes his actions in the Bayesian way. However, given that such a rationalization always exists, it seems a bit pedantic to insist that a decision maker should actually provide it.

2.4 Bayesian Replies

This line of criticism usually meets four replies. First, it is often said that Bayesianism provides the definition of rationality. Thus, proceeding along Bayesian lines is rational by definition, and that is all there is to say. Second, Bayesians often cite the old adage 'garbage in, garbage out', meaning that, of course, we get absurd results from the Bayesian machinery when we feed it with absurd prior probabilities. Third, with respect to decision making, it is said that the decisions taken by a Bayesian decision maker are optimal in the decision maker's eyes; the fact that another prior probability distribution would lead to other decisions is not a valid criticism from the decision maker's point of view. Fourth, Bayesians provide examples where the Bayesian calculus actually excludes some possibilities or leads to definitive results.

All four replies fail to answer the criticism raised above. First, Bayesianism was meant as an answer to irrationalism, that is, as a solution to a specific problem. I have argued that it is just a restatement of irrationalism in probabilistic terms. How Bayesians would like to use the word 'rationality' is completely irrelevant for this, or any other, substantive issue.

Second, 'garbage in, garbage out' also misses the point. Irrationalism says that there is no difference between garbage and no garbage. Bayesianism was

[8] For the anything-goes theorem, see Albert 2001; 2003 and, for a formal proof, 1999, theorem 1. Theorem 2 in Albert 1999 is mistaken and should be ignored.

supposed to tell the garbage from the rest. If the Bayesian calculus needs to be fed the solution to the problem it was meant to solve, it is a failure.

Third, it is irrelevant for the evaluation of Bayesianism whether Bayesians are convinced that their decisions are optimal. After all, non-Bayesians may also be convinced that their decisions are optimal, but Bayesians would not count these convictions as arguments against Bayesianism.

Fourth, the fact that the Bayesian calculus seems to have some bite in statistics and decision making is simply due to the fact that, in these applications, only a few prior probability distributions are taken into consideration. This is done mainly by assuming that a lot of things are impossible. For instance, the Bayesian calculus immediately leads to sharp conclusions once we set $p_{00} = 0$ and $p_{11} = 0$ in figure 1. This is just the point of the anything-goes theorem: any conclusions result from the choice of the prior probability distribution, but Bayesianism does not help us in choosing this distribution.

Let me focus on one argument that, historically, has been an especially important selling point for Bayesianism: the Dutch book argument.[9] This argument considers a decision maker who must, for all possible bets on events, state his subjectively fair odds, that is, the odds at which he would take either side of the bet. A bookmaker is then allowed to select any number of bets, the side the decision maker has to take, and the (monetary) stakes. Then, the selected bets are played out and the decision maker wins or loses money depending on the course of events. How should the decision maker choose the odds?

Since the Dutch book argument concerns only probabilities, we can for the moment assume that the decision maker is risk neutral and interested only in the monetary payoffs from his bets. For a Bayesian, the subjectively fair odds must then be equal to the ratio of subjective probabilities. For instance, in our numerical example, the fair odds for a bet on $X = 0$ are $P(X = 0)/P(X \neq 0) = 0.60/0.40 = 3{:}2$ since $X \neq 0$ means $X = 1$. In a bet on $X = 0$ with odds 3:2, where the winner gets S (the stakes), the person betting on $X = 0$ contributes $0.60S$ and the person betting against $X = 0$ contributes $0.40S$.

Let us call the set of odds coherent if all the odds are based in this way on a single subjective probability distribution. If and only if a decision maker's odds are incoherent, a bookmaker can pick some bets in such a way that the decision maker will lose money no matter what happens. Such a set of bets is called a Dutch book. For instance, a simple case of incoherent odds is 3:2 for $X = 0$ and 3:2 for $X = 1$. If the decision maker states these odds, the bookmaker takes the less favored side of both bets at equal stakes S. This is a Dutch book. The decision maker pays two times $0.60S$ and wins one of the bets, receiving S.

[9] The Dutch book argument is a probabilistic version of the money-pump argument. It can be used, less ambitiously, to argue that numerical degrees of belief used as decision weights should obey the axioms of the probability calculus; see Hájek 2009 for a survey of the literature. My criticism of the argument, however, is that it has no force if one rejects the assumption that degrees of belief are necessary, as, e.g., critical rationalits and classical statisticians do. The modifications of the Dutch book argument discussed by Hájek are irrelevant to this line of criticism; therefore, I present the simplest version of the argument.

This leaves the decision maker with a loss of $0.2S$ and the bookmaker with a corresponding profit.

Betting on the basis of a prior probability distribution, then, may protect against accepting a Dutch book. While this is correct as far as it goes, it is hardly convincing as an argument in favor of Bayesianism. First of all, the Dutch book argument assumes a situation where some option (accepting a Dutch book) is worse than others (for instance, not betting), no matter what happens. In order to avoid such strictly dominated options, no subjective probabilities are needed. Thus, non-Bayesians would also reject Dutch books. Second, and crucially, the problem of induction implies that there are no strictly dominated options; strict dominance requires that we have already excluded certain things as impossible. Therefore, the Bayesian promise of protection is empty.[10]

3. The Dilemma of Perfect Rationality

In economics, two different conceptions of rationality are used: perfect rationality and bounded rationality. The usual exposition goes like this (Simon 1987; Gigerenzer and Selten 2001; Selten 2001). Perfect rationality, which is identified with Bayesian rationality, is the practically unachievable but theoretically and normatively important ideal version of rationality. Bounded rationality, in contrast, is the rationality of actual human decision making. It is often identified with rule-following behavior, where the rules (or heuristics) may perform well in some situations but rather badly in others. As the name already suggests, bounded rationality is viewed as inferior to perfect rationality. Boundedly rational agents intend to be rational but fail; they are constrained by their limited cognitive capacities. In order to be perfectly rational, they would need superhuman intellectual powers: infallible and complete memory, lightning speed and perfect accuracy in complicated calculations, and so on.[11]

In this exposition, one crucial ingredient is missing. What, exactly, are the hypothetical advantages of being perfectly rational? Intellectual superpowers are just the inputs needed for achieving perfection. What is the value added of perfect rationality?

In the last section, I have argued that there is no value added if perfect rationality is identified with Bayesianism. Is there any other candidate for perfect rationality that would do a better job? This is hard to tell since it is unclear what perfect rationality is meant to achieve. However, arguments like the Dutch book

[10] The Dutch book argument creates an illusion of certainty by excluding the actual betting process from the set of possible worlds. If, however, betting behavior is included in the set of possible worlds, it cannot be ruled out that events are influenced by the parties' betting behavior. A trivial example: The decision maker may accept a Dutch book with a certain loss of one thousand euros because he believes that this will make his rich uncle Bill take pity on him and send a gift of one million.

[11] Gigerenzer and Selten (2001a, 9) seem to oppose the standard view that perfect rationality is better in some sense than bounded rationality. In the same volume, however, Selten (2001, 13–15) views Bayesian rationality as the (unachievable) ideal.

argument and discussions comparing perfect and bounded rationality (Gigerenzer 2001, 40–43) suggest two criteria as necessary for perfection: being perfectly rational should help us to achieve our goals, and the principles of perfect rationality should be independent of the environment.

Once the two criteria are on the table, however, it becomes rather obvious that perfect rationality does not exist. The argument can be stated as follows.

Premise 1 (usefulness):
If it exists, perfect rationality is useful in all possible worlds.
Premise 2 (apriorism):
If it exists, perfect rationality is the same in all possible worlds.
Conclusion:
Perfect rationality does not exist.

Of course, the premises are somewhat imprecise, but they should suffice even in the present form to convey the message. In order to be useful, perfect rationality must say more than just 'anything goes'; it must restrict decision making such that being perfectly rational is advantageous. However, for any kind of restriction, we can imagine some world where a given restriction on decision making is disadvantageous. Hence, usefulness and apriorism are inconsistent, and perfect rationality does not exist.

The argument above, then, gives rise to a dilemma. Any conception of rationality must give up either usefulness or apriorism. I have shown in the preceding section that Bayesianism satisfies apriorism at the price of uselessness. In the next section, I explore the alternative of giving up apriorism.

4. From Ecological to Critical Rationality

4.1 Belief Formation as a Decision Problem

When we consider belief formation as a decision problem, we have to state a goal first. The goal in rational belief formation is to accept (that is, believe) true statements and to reject false statements (which means to accept the negation of false statements, which are true). A naturalistic conception of rational belief formation requires that we use decision rules that we can rationally believe to be effective in achieving this goal. Or rather, since a comparative judgment seems to suffice: decision rules that we can rationally believe to be at least as effective as any of their competitors.

There are several rules that are typically accepted as rules of rational belief formation. First, it is rational to believe the deductive consequences of rational beliefs. Second, it is rational to believe what one observes (observational beliefs), at least under certain circumstances. We will consider only conceptions of rationality that accept these rules; for this reason, we skip any discussion of them.[12]

[12] Deductive logic as well as observational beliefs can be criticized. In fact, deductive logic as we know it has developed over time as a response to criticism. On the criticism of observational

Classical inductivism, the philosophical precursor of Bayesianism, assumes that it is rational to believe in non-deductive, or inductive, consequences of rational (specifically, observational) beliefs. However, non-deductive arguments are equivalent to deductive arguments with additional premises. In the context of inductivism, these additional premises are called inductive principles. Given that inductive principles are not themselves inductive or deductive consequences of rational beliefs, classical inductivism is necessarily incomplete: it contains no rule that allows for rational belief in inductive principles. Consequently, it adds nothing to the first two rules.

The first two rules, however, are not sufficient for solving the problem of induction. According to them, it is never rational to believe any proposition about future events or, more generally, events that have not yet been observed. Hence, if one restricts rational belief formation to the application of these two rules, irrationalism cannot be avoided. A different conclusion requires adding at least one further rule. This rule must be useful and it must recommend itself: it should be rational to be rational.

4.2 Ecological Rationality

Useful principles of rationality cannot be useful in all possible worlds; they must be adapted to the world we actually live in. Adaptation to the environment features prominently in "ecological rationality" (Gigerenzer et al. 1999; 2001; Todd and Gigerenzer 2000, 2007). This conception of rationality is based on a specific theory of decision making, the adaptive-toolbox theory, which is a version of bounded rationality.

According to the adaptive-toolbox theory, decision makers use decision heuristics that are composed of cognitive and emotional building blocks. The building blocks form an adaptive toolbox: they can be recombined to form new heuristics that are adapted to new tasks. Heuristics are domain-specific and can take advantage of the actual structure of their domain of application. If they are adapted to the environment in which they are used, they can be successful as well as 'fast and frugal', that is, they can work well within the limits of human cognitive capacities. Of course, heuristics are no algorithms; they are not guaranteed to find a correct solution to a problem. Nevertheless, they may work better than alternatives like random decision making.

Consider the following example of a simple decision task (Todd and Gigerenzer 2000, 732–733). Experimental subjects are asked, for several pairs of cities, to decide which of the two has more inhabitants. Typically, subjects do not know the answers; they have to guess. However, if they recognize the name of only one city in a pair, more likely than not the name they recognize is the name of the larger city. This holds because the names of larger cities come up more often in conversations or the news, which makes it more likely that subjects recognize these names.

beliefs from the standpoint of criticial rationalism, see Andersson 1994; and Musgrave 1999, 341–347.

Thus, choosing the option one recognizes is a good and intellectually undemanding heuristic in this decision problem. It is, however, domain specific; it cannot even be applied to problems where all options or no options are recognized. Moreover, in order to use this heuristic, subjects need to know that name recognition is correlated with size. If they blindly decided, in every kind of comparison between two cities, for the city whose name they recognize, one could easily trick them by asking after the smaller city or the city with more cows per head of population.

The adaptive-toolbox theory is not at issue here. What concerns us is the conception of ecological rationality. Their proponents describe it in the following words:

> "The 'rationality' of domain-specific heuristics is not in optimization, omniscience, or consistency. Their success (and failure) is in their degree of adaptation to the structure of environments, both physical and social. The study of the match between heuristics and environmental structures is the study of ecological rationality." (Gigerenzer 2001, 38)

> "[B]ecause the human mind has been shaped by the adaptive processes of evolution and learning, we predict that people will tend to be ecologically rational themselves, often using simple decision heuristics that confer the twin advantages of speed and accuracy in particular environments." (Todd and Gigerenzer 2007, 169)

However, these passages state only some requirements for a conception of rationality, while the city-size example—together with other examples—illustrates only the basic idea of heuristics adapted to the environment. The organization of the adaptive toolbox, and the crucial point of rational belief formation in particular, is left in the dark.

Of course, the city-size example considers the formation of beliefs about the relative size of cities. However, it is assumed that subjects already know that city size is correlated with name recognition. This known correlation serves as an inductive principle, transforming belief formation into a simple case of deduction: given the correlation and the fact that only one name was recognized, it follows deductively that it is more probable than not that this name belongs to the larger city. However, as has already been explained, the problem of rational belief formation cannot be solved by assuming given inductive principles.

The city-size example, then, leaves open the question of how to find the right heuristic (or the right inductive princple). For choosing between heuristics, Todd and Gigerenzer (2000, 771) propose a further layer of meta-heuristics build from the same components as the lower-level heuristics. This, of course, invites the question of how to choose between meta-heuristics. Todd and Gigerenzer stop the incipient regress of heuristics, conjecturing that meta-heuristics are tried in the order of past successfulness, where the ordering is provided by a process of reinforcement learning.

From this description, it remains unclear how much is assumed to be known by the decision maker. Anyway, the highest level of the adaptive toolbox is a process of reinforcement learning. While reinforcement learning may describe some aspects of actual human decision making, it is certainly not a candidate for the decision rule we are searching for. Even if one could argue that it is rational to learn in this way, reinforcement learning cannot recommend itself. As a conception of rationality, ecological rationality is too simple.

4.3 Critical Rationality

In contrast to ecological rationality, critical rationalism explicitly proposes a third decision rule for rational belief formation: it is rational to believe a hypothesis if it has so far withstood serious criticism better than its competitors.[13]

Criticism proceeds by deduction. The most prominent case of criticism is falsification, where a hypothesis is shown to be inconsistent with observational beliefs. In the case of a falsification, a hypothesis is rejected independently of the fate of its competitors because a falsification means that a hypothesis has succumbed to criticism. The same holds in the case where it is shown that a theory is contradictory. In the case of metaphysical hypotheses, criticism may be comparative; for instance, if all competing hypotheses have the same unfortunate consequence, this would usually not count against any of them. Much of the content of critical rationality lies in the various lower-level principles of how to criticize beliefs, including various domain-specific scientific methodologies.[14]

The difference between critical rationalism and classical inductivism is that critical rationalism allows us to believe in a hypothesis without providing some deductive or non-deductive argument with the hypothesis as a conclusion. There exists a deductive argument with the conclusion that it is rational to believe in the hypothesis, but this conclusion is a statement about the hypothesis, not the hypothesis itself. As Musgrave (1999) argues, the deductive argument is an argument in favor of believing the hypothesis—an act or a decision—, not an argument for the hypothesis. This makes a big difference. First, since it is a deductive argument, there is no inductive principle or other missing premise whose rational acceptability must be shown. Second, since the conclusion is not

[13] The following summary of critical rationalism is based on the much more detailed account by Musgrave 1993, ch. 15; 1999. Specifically, for the full version of critical rationalism, which takes observational beliefs and epistemic division of labor into account, see Musgrave 1999, 347–350. Some critical rationalists disagree with Musgrave, most prominently Miller 1994; on this debate, see Musgrave 1999, 332–336.

[14] In my view, critical rationality can also be viewed as defining a process for reaching, by criticism, a reflective equilibrium where all beliefs are rational. There are several other versions of the idea of a reflective equilibrium: Rawls' and Goodman's version are discussed by Hahn 2004. I would add the Savage-Binmore variant of Bayesianism where one should massage the prior until one can accept its hypothetical consequences (Binmore 2009). All these other ideas of reflective equilibrium seem to lack a good description of the dynamics leading to equilibrium. See, in contrast, Kliemt 2009, 37–41; Kliemt uses theory absorption as a critical argument in order to derive the Nash solution as a reflective equilibrium in game theory.

the hypothesis itself, there is no presumption that the premises already imply the hypothesis.

When we say that it is rational to believe a hypothesis that so far has withstood serious criticism better than any competitor, we claim that this decision rule is at least as effective in accepting true hypotheses and rejecting false hypotheses as any competing decision rule. To believe this claim is rational because critical rationality has withstood serious criticism better than competing rules. Thus, critical rationality recommends itself. Yet, it does not immunize itself against criticism. It could conceivably be criticized and rejected on its own terms.

Let us assume that critical rationalists are right and critical rationality is a useful and comprehensive solution to the problem of rational belief formation. We can, of course, easily imagine possible worlds in which critical rationality would not be useful, for instance, just to mention one of many logical possibilities, a world with gods that favor uncritical belief over critical thinking. Usefulness implies that critical rationality exploits some features of the actual world, for instance, that such gods do not exist in the actual world.

If we knew all the features of the actual world exploited by critical rationality, we could deduce from this knowledge the usefulness of critical rationalism. We would, then, be in a situation analogous to the city-size example: our knowledge would provide us with an inductive principle. However, under our assumptions, this inductive principle must be equivalent to critical rationality; thus, it does not matter that we do not know it. Moreover, there is no way we could be sure that the inductive principle is true—as we cannot be sure that there is nothing better than critical rationality. Anyway, speculation about this unknown inductive principle is not helpful. All we have is, on the one hand, some inductive principles proposed so far, which have not withstood criticism, and critical rationality, which has done so.

5. Conclusion

The extension of critical rationality from belief formation to choice of actions is simple in principle, although the details are not easily worked out.[15] The general idea is that a proposal for the solution of a decision problem is an implicit hypothesis, namely, that implementing the proposal achieves the decision maker's goal. It is rational, then, to implement a proposal if and only if it is rational to believe in the corresponding hypothesis.

In practical decision-making, different solution proposals to the same problem may be rationally implemented since there may be different courses of action achieving the given goals. When several proposals survive criticism equally

[15] See Miller 2009 for a discussion of the relevant literature and many important aspects. Miller's general perspective is, however, more complicated because he presupposes a much weaker version of critical rationalism.

well, no rational decision between them is possible. It is also possible that no proposal survives criticism, in which case no rational decision exists.

The case where no rational decision exists is a case where the goals may be criticized. Since people are not transparent to themselves, goals are hypothetical and criticizable, at least on account of their unattainability or their incompatibility.[16]

Critical rationalism shows more explicit continuity with evolutionary theory than ecological rationality. It is firmly embedded in evolutionary epistemology, which emphasizes the continuity between all kinds of cognitive processes including creativity in science and elsewhere (see Bradie and Harms 2008 for a survey). This leads to a slightly different emphasis in decision theory. For instance, the choice between decision heuristics can be viewed as a process where several proposals are generated blindly and then criticized until some proposal survives criticism or time runs out. This is a continuation of the evolutionary process, where organisms (or genes) can be viewed as embodying hypotheses or decision rules that are 'criticized' by nature whenever they fail to reproduce (or to reproduce at a higher rate than their competitors).[17]

Critical rationality, then, is not only relevant in the philosophy of science. It should also be viewed as the solution to the problem of rational decision-making and replace Bayesianism, which is not satisfactory in either field. Critical rationality shares the emphasis on evolutionary aspects with ecological rationality, which has been proposed as an alternative to Bayesianism, or perfect rationality, in economics. In contrast to critical rationality, however, ecological rationality cannot serve as a comprehensive conception of rationality.

References

Albert, M. (1999), "Bayesian Learning when Chaos Looms Large", *Economics Letters* 65, 1–7.

— (2001), "Bayesian Learning and Expectations Formation: Anything Goes", in: Corfield and Williamson (2001), 341–362.

— (2003), "Bayesian Rationality and Decision Making. A Critical Review", *Analyse & Kritik* 25, 101–117.

Andersson, G. (1994), *Criticism and the History of Science. Kuhn's, Lakatos's and Feyerabend's Criticisms of Critical Rationalism*, Leiden: Brill.

Bartley, W. W., III (1987), "Theories of Rationality", in: Radnitzky and Bartley (1987), 205–214.

Binmore, K. (2009), *Rational Decisions*, Princeton–Oxford: Princeton University Press.

[16] At this point, it might again be useful to distinguish, with Kliemt 2009, between goals and preferences, where the latter are viewed as implicit hypotheses concerning the best way to trade off different goals against each other.

[17] Todd and Gigerenzer never mention evolutionary epistemology but refer to Simon, who has been criticized by Campbell (1960, 1974) from an evolutionary perspective.

Bradie, M. and W. Harms (2008), "Evolutionary Epistemology", in: Zalta (2009), URL: .../entries/epistemology-evolutionary/.

Campbell, D. T. (1960), "Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes", repr. in: Radnitzky and Bartley (1987), 91–114.

— (1974), "Evolutionary Epistemology", repr. in: Radnitzky and Bartley (1987), 47–89.

Corfield, D. and J. Williamson (2001) (eds.), *Foundations of Bayesianism*, Dordrecht–Boston–London: Kluwer.

Earman, J. (1992), *Bayes or Bust?*, Cambridge/MA: MIT Press.

Gigerenzer, G. (2001), "The Adaptive Toolbox", in: Gigerenzer and Selten (2001a), 37–50.

— and R. Selten (2001), "Rethinking Rationality", in: Gigerenzer and Selten (2001a), 2–12.

— and — (2001a) (eds.), *Bounded Rationality. The Adaptive Toolbox*, Cambridge/MA–London: MIT Press.

Gillies, D. (2001), "Bayesianism and the Fixity of the Theoretical Framework", in: Corfield and Williamson (2001), 363–379.

Hahn, S. (2004), "Reflective Equilibrium—Method or Metaphor of Justification?", in: W. Löffler and P. Weingartner (eds.), *Knowledge and Belief—Wissen und Glauben*, Proceedings of the 26th International Wittgenstein Symposium, Wien: öbv & hpt, 237–243.

Hájek, A. (2009), "Dutch Book Arguments", in: P. Anand, P. Pattanaik and C. Puppe (eds.), *Oxford Handbook of Rational and Social Choice*, Oxford: Oxford University Press, 173–196.

Howson, C. and P. Urbach (1993), *Scientific Reasoning. The Bayesian Approach*, 2nd ed., La Salle: Open Court.

Kliemt, H. (2009), *Philosophy and Economics I: Methods and Models*, München: Oldenbourg.

Miller, D. (1994), *Critical Rationalism. A Restatement and Defense*, Chicago—La Salle: Open Court.

— (2009), "Deductivist Decision Making", University of Warwick, URL: http://www2.warwick.ac.uk/fac/soc/philosophy/people/associates/miller/kpf.pdf.

Musgrave, A. (1993), *Common Sense, Science, and Scepticism. A Historical Introduction to the Theory of Knowledge*, Cambridge: Cambridge University Press.

— (1999), "Critical Rationalism", in: A. Musgrave, *Essays on Realism and Rationalism*, Amsterdam–Atlanta: Rodopi, 314–350.

Radnitzky, G. and W. W. Bartley, III (1987) (eds.), *Evolutionary Epistemology, Theory of Rationality, and the Sociology of Knowledge*, La Salle: Open Court.

Selten, R. (2001), "What is Bounded Rationality?", in: Gigerenzer and Selten (2001a), 13–36.

Simon, H. A. (1987), "Bounded Rationality", repr. in: J. Eatwell, M. Milgate and P. Newman (1990) (eds.), *The New Palgrave: Utility and Probability*, New York–London: Norton, 15–18.

Wallace, R. J. (2009), "Practical Reason", in: Zalta (2009), URL: ../entries/practical-reason/.

Zalta, E. N. (2009) (ed.), *The Stanford Encyclopedia of Philosophy*, summer 2009 ed., URL: http://plato.stanford.edu/archives/sum2009.