be deserved; I’m inclined to think that it can be. It is because punishment is not deserved just because one has done moral wrong; for one can innocently (that is, inculpably) do wrong. It is not wrongdoers as such who deserve punishment; it is (if anyone) those who are culpable who deserve it (see Distinction 3). Sher might respond that it is culpable wrongdoers that he had in mind all along (although this is not what he says; he writes simply of wrongdoers). But, if so, then his position faces a further problem, even if all the points made hitherto can be satisfactorily handled. The problem is that it is not degree of wrongness but degree of culpability that must determine the severity of retributive punishment. No clue is given as to how this might be measured, nor is any clue given as to how punishment might be tailored to fit degree of culpability rather than degree of wrongdoing. In the absence of such an account, the notion of just punishment is left wholly unelucidated.\(^\text{11}\)

**Notes**


4See also *Collective and Corporate Responsibility*, pp.165–6.

5It is worth noting that Dennett himself regards consciousness, indeed self-consciousness, as necessary for moral responsibility.

6As he does say on pp.188–90 of *Collective and Corporate Responsibility*.

7Perhaps he is in some way morally on a par with someone who has not luckily avoided inflicting further harm, but that is another matter.


11My thanks to Peter French and Terry McConnell for comments on earlier drafts.

Robert Nozick, *The Nature of Rationality*\(^1\)

David Christensen

University of Vermont

1. Introduction

One of the things that makes philosophical conversation valuable is the relatively free flow of ideas it can promote. Released from some of the strictures of argument that attend so much contemporary written philosophy, participants in a philosophical conversation are able to conjecture rather than prove, to float ideas without first worrying over their
ultimate defensibility. This lack of constraint not only allows for a certain kind of philosophical fun; it also fosters an intellectual creativity that is central to philosophy’s serious concerns.

Something of the conversational spirit animates much of Robert Nozick’s *The Nature of Rationality*. The prose is informal, sprinkled with parenthetical asides and with questions, some of which are used rhetorically, but many of which simply invite readers to wonder about issues Nozick wonders about. Even substantive suggestions are sometimes made without detailed considerations of arguments for and against them. One result of this is that a surprising number of bold ideas have been put forward in a relatively short and accessible book.

Nozick’s topic encompasses various aspects of rationality, from rationality in decision to rationality in belief and even in preferences. Although it would be impossible to discuss here all of the ideas put forward in the book, I would like to examine some of its most important suggestions. I will turn first to a fairly radical pair of proposals Nozick makes for revising the formal theory of rational decision.

### 2. Evidential vs. Causal Decision Theory

In standard evidential treatments of decision theory, a rational agent is said to maximize what Nozick calls the Evidentially Expected Utility (EEU) of her act. This quantity is calculated by adding up the utility of the act’s various possible outcomes, with the utility of each outcome weighted by the probability of that outcome *on the condition that the agent performs the act in question*. The reason that the probabilities of outcomes are conditional on the act in question is that the probabilities of the outcomes are often affected by the agent’s choice of act.

Sometimes, however, the way that the choice of a certain act affects the probability of an outcome is not due to any way in which the choice of that act would influence whether that outcome would occur. Sometimes these conditional probabilities reflect the fact that a certain act would be *evidence* about whether the outcome would occur. Such is the case in Newcomb’s problem, which was made famous by an early paper of Nozick’s (see his (1969)).

Imagine that you are confronted with the following situation. You are seated in front of two boxes. In the clear box, you can see $1,000. In the opaque box, there is either $1,000,000 or nothing. You are to choose between taking just what is now in the opaque box, or taking what is now in both boxes. Sounds easy enough—surely anyone but an idiot would take both! But then you are told the following additional fact: that an extremely reliable Predictor has studied you, and has put the $1,000,000 in the opaque box if and only if he has predicted that you will take the contents of the opaque box only.

Now some would say—and I should own up here to being one of them—that this new information should not affect your decision at all. The Predictor has made his choice, the money is either there or it isn’t, and you might as well take the extra $1,000. But not everyone sees things this way, and by using evidential decision theory, one can see why. The reliability of the predictor makes the probability of my getting rich, on the condition that I take both boxes, extremely low. On the other hand, the probability of my getting rich on the condition that I take just one box is very high. Thus the EEU of taking both boxes is much lower than the EEU of taking just one!

A sizable literature has grown up around this problem. Some advocates of the two-box solution see the example as a *reductio* of the EEU approach to decision, and have
formulated sophisticated theories of decision designed to separate evidential from causal aspects of one’s acts. (These accounts have rational agents maximize a quantity Nozick calls Causally Expected Utility, or CEU.) Other two-box advocates have argued that, with appropriate assumptions or emendations, the EEU-based account actually recommends taking both boxes. On the other side, advocates of the one-box solution have pointed to the fact that if people were actually given a chance to play the Newcomb game, the one-boxers would end up a lot richer. They see no need for revising evidential decision theory.

Nozick believes that neither side has provided a completely satisfactory resolution of the problem. He proposes to make progress by way of considering an aspect that has been little discussed in the literature. Nozick examines Newcomb cases in which the amount of money in the transparent box is much less, or much greater, than $1,000. If there is only a very small amount of money in the transparent box, Nozick claims that two-boxers will not be willing to take both, showing that they are somewhat moved, after all, by the EEU argument. Conversely, Nozick claims that one-boxers will switch to the two-box strategy when the money in the transparent box increases dramatically, even if the increase is not enough to make the EEU of taking two boxes higher than that of taking only one.

Nozick draws two conclusions from his claim that people’s intuitions in the Newcomb case are sensitive in this way to the amount of money in the transparent box. The first is that “no one has complete confidence in the argument he or she follows for Newcomb’s initial example” (45). We all harbor some sympathy for both EEU and CEU considerations. To describe this attitude, Nozick suggests formulating a decision theory which has us maximize a weighted sum of the CEU and EEU for an action (Nozick calls this quantity the Decision Value (DV) of the act). Nozick’s second, and more provocative, conclusion is that it is DV, rather than EEU or CEU, that provides a normatively correct principle of choice.²

Nozick notes that, unlike EEU-based or CEU-based accounts, a DV-based decision theory can rationalize switching between one-box and two-box answers to versions of Newcomb’s problem involving differing amounts of money in the transparent box. But Nozick’s suggestion, if correct, would go well beyond providing an answer to Newcomb’s problem; it would constitute a major development in decision theory. Thus it is worth looking carefully at the reasons that are being given for this reformulation.

To begin with, it is not at all clear that the intuitions that we should switch between one-box and two-box answers are nearly as universally shared as Nozick claims. I, for one, would be happy to take both boxes even if the transparent one were empty. (After all, taking the second box would cost me no money, and I would gain the satisfaction of not bowing to irrational impulse!) Perhaps I am unusually dogmatic, but conversations with colleagues revealed only one out of four who favored switching.³

But let us put this issue aside. Nozick may have done a more substantial survey, and may have posed the problem in a more neutral way, than I did. Suppose that most people do have intuitions in Newcomb problems that are sensitive to the amount of money in the transparent box. How seriously should we take such a fact?

One thing that seems quite clear is that these sorts of intuitions should not be decisive in themselves. Ordinary intuitions about hard cases—even strong, widely shared intuitions—are oftentimes misguided. Russell’s paradox reveals that logical contradictions may lie in intuitively innocuous thoughts. And misguided intuitions seem especially prevalent in cases involving probabilities. Members of a moderate size undergraduate class will typically vastly underestimate the probability that at least two class members share a birthday. The gambler’s fallacy is widespread, and a number of other widely shared statistical
fallacies have been studied by cognitive psychologists (see Kahneman, et al. (1982)). And even professional philosophers and mathematicians can have trouble with some cases, such as the “three prisoners problem” mentioned by Nozick (or its recently illustrious cousin, the “Monty Hall problem”). Sometimes the misguided intuitions disappear when one comes to understand a situation clearly. But frequently even those who understand the sources of these illusions continue to feel their pull.

What is really needed to solve the Newcomb problem, then, is much more than agreement with a certain set of intuitions about cases. What’s more important is doing justice to philosophical considerations on each side. There is a substantial and sometimes subtle literature addressed to these considerations. Nozick provides references to this work, and even reports on Howard Sobel’s reminder to him that causal theorists have offered not only arguments for their favored principle, but diagnoses of how mistaken intuitions on the other side arise. Nozick himself saw this as important, and gave interesting diagnoses of one-boxers’ intuitions, in his original paper. Unfortunately, however, the body of argument developed in this literature is simply not confronted in *The Nature of Rationality*.

Part of the reason for this may be that Nozick sees the literature as not having produced the *kind* of argument on either side that would be required to justify confidence in its favored principle of decision. He writes:

> It is somewhat strange that writers on decision theory generally have shown such confidence in their views. For if we formulate the issue about the correct principle of decision as a decision problem, one about which principle of decision should be followed—we might imagine that pills have been developed that can transform us into consistent followers of each principle—then it is not obvious what the contending principles of decision will answer. (47)

Nozick notes that for inhabitants of certain kinds of worlds (ones with a lot of Newcomb situations, for instance) both the EEU and CEU-based decision theories might recommend taking a pill that made one into a EEU-maximizer. Similarly, for other types of worlds, both decision-theories would recommend taking the CEU-maximizer pill. Presumably, in some sorts of world (though Nozick offers no evidence that ours is one of them), the best pill to take would be universally acknowledged to be one that made us give certain intermediate weights to both EEU and CEU.

Now the question Nozick is raising here—what happens when we treat the choice of decision theory as a decision problem in itself?—is an intriguing one. Moreover, quite a bit of recent work in epistemology has been sympathetic to the suggestion that what epistemic rules are rational to follow depends on the character of the actual world, a suggestion that would seem just as plausible when applied to decision rules. So Nozick’s challenge is, at the methodological level, extremely interesting. But we should be very careful about the more specific suggestion: that one can answer the question of which theory describes rational decision by formulating the question as one of which pill it would be rational to take in Nozick’s thought-experiment.

In thinking about the significance of this thought-experiment, we must be careful to distinguish between the rationality of adopting a decision-policy (perhaps by taking a pill), and the rationality of particular decisions made in accord with that policy. Suppose that the Thought Police routinely detected and tortured those who were predisposed to account for
their own biases when making decisions. In such a world, it might be eminently rational to
take a pill that would make one predisposed to make decisions without accounting for
one’s biases. But would the decisions one reached after taking the pill be rational ones?
The same issue arises in the Newcomb case. The fact that a perverse and powerful
predictor would punish one for being predisposed to maximize CEU might make it
rational to take the EEU-maximizer pill. But this does not show that the EEU-based
decisions one would then make would be rational decisions.

One might reply that if anything makes a decision rule rational, it is that being a
follower of that rule promotes the agent’s ends. And, as advocates of EEU-maximization
point out, one-boxers fare much better, financially speaking, in Newcomb situations. In a
world full of Newcomb situations, wouldn’t the flourishing of the one-boxers render
EEU-maximizing rational?

It would be a mistake, I think, to argue so quickly from the flourishing of the EEU-
maximizers to the rationality of the EEU-maximizing decision rule. For there are at least
two ways in which being a follower of a certain rule could be beneficial. One is that the
decisions made in accordance with the rule could tend to be beneficial. The other is that
the agent could benefit from being a follower of the rule in some collateral way—a way
independent of the results of the decisions made in accordance with the rule.

With this distinction in mind, it should be remembered that one-boxers do not fare
better as a result of their decisions. They fare better because the Predictor predicts that
they will make certain decisions. True, these predictions are in turn presumably explained
by the one-boxers’ EEU-maximizing tendencies. But calling a decision rule rational in
virtue of this sort of collateral benefit accruing to the rule’s followers is not nearly as
plausible as calling a rule rational in virtue of benefits conferred by the decisions made in
accordance with the rule. The first sort of benefit might help render a rule rational to
adopt. But this is not the same as helping to render the decisions made in accord with
the rule rational. Nor, I would argue, is it the same as helping to render the rule a rule of
rationality.

Thus Nozick poses an interesting question about choice and validation of theories of
rational decision. But, to my mind, the arguments put forth in The Nature of Rationality
in favor of Nozick’s “split the difference” settlement of the EEU/CEU controversy are not at
all persuasive. (My suspicion is that this sort of compromise will satisfy fewer people than
would either causal or evidential decision theory in pure form.)

But let us leave this issue. Nozick intends his notion of Decision Value to do more than
settle the causal/evidential controversy. There is another dimension of our acts (beyond
EEU and CEU) that Nozick wants to encompass in his full account of DV. The idea that
we need to add this new dimension to our evaluation of acts is itself a provocative and
independently motivated claim that would make an enormous change in decision theory. It
is to this claim that I turn in the next section.

3. Symbolic Value

An act may cause a situation to occur, and an act may be evidence that a situation will
occur, but there is a third relation between acts and situations that Nozick would accord an
important place in decision theory: an act may symbolize a situation. Nozick notes that
some actions whose explanations rely heavily on symbolic aspects of acts are paradigmatic-
ically irrational; compulsive handwashing provides a plausible example. But such exam-
people do not begin to account for the importance of symbolism in our lives, and many other
texts are much more obviously attractive. Acting in accordance with one’s moral values is symbolically valuable in aligning oneself with the good and the righteous.
(Nozick does not want to assimilate moral value to symbolism; rather he holds that there is
symbolic value to moral acts in addition to, and because of, their moral value.) And one
can easily think of countless more cases, from sending a loved one a rose to wearing an
AIDS awareness ribbon to eating special foods on holidays, in which symbolic meanings
give us reasons for action.

Nozick points to some interesting ways in which symbolic considerations contribute to
intuitively reasonable behavior. Consider the case of someone facing a temptation to break
his diet. The utility of eating the well-buttered popcorn now looms large due to its
immediacy. But if the agent sees eating the popcorn now as standing for or symbolizing
the general class of diet-breakings, it will be harder for him to think of himself as cheating
“just this once.” Symbolic utility thus helps us be principled in our actions.

Nozick also wants to use symbolic utility to help provide reasons for pursuing cooperate
strategies in Prisoner’s Dilemma situations. (The typical example here involves a pair
of prisoners, held incommunicado, each told that his prison term will be reduced if he
defects and informs on the other, but that the terms they receive if they both defect will be
longer than the terms they will receive if neither defects. What is interesting about such
situations is that each prisoner, considering only his self-interest, can calculate that no
matter what the other does, he will be better off defecting; nevertheless, if they both
pursue this course of action, each will be worse off than if they both refused to defect.)
Nozick points out that the symbolic value of being a cooperative person could lead
someone who acted in part so as to maximize symbolic utility to adopt the cooperative
(i.e. non-defecting) approach in Prisoner’s Dilemma-like situations.4

Now Nozick is undoubtedly right in holding that a satisfactory theory of rational
decision must take account of the fact that symbolic aspects of acts frequently provide us
with reasons for action. The key question this raises for decision theory is this: can the
symbolic utility of acts be represented adequately in a standard CEU-maximizing or EEU-
maximizing theory, or does it require some sort of structural change? After all, many
different sources of value are already collected together in the standard accounts. The
various hedonistic values people derive from omelettes, appropriate wines, smoking, and
sex are obviously included, but so are the perhaps more complicated benefits people
derive from music, career choices, and personal relationships, as well as values people
find in moral or religious aspects of situations. Utility—as usually conceived—flows
from preferences, and the objects of our preferences are many and various. Mightn’t the
symbolic value of acts be best thought of as just one amongst these? Nozick contends that
it is not. Thus his revised formulation of his notion of Decision Value looks like this:

\[ DV(A) = Wc \times CEU(A) + We \times EEU(A) + Ws \times SU(A) \]

where SU is Symbolic Utility (the W-factors are the weights assigned to the act’s causal,
epistemic and symbolic utilities).

I can discern three arguments in The Nature of Rationality to show that symbolic value
must be treated as a quantity separate from other sources of utility. One is that “we might
want to keep track of symbolic utility, since we think it appropriate to give this factor
different weight in different choice situations” (48). But although it seems true that the
importance one gives to, say, symbolizing one’s love for another person will vary in
different situations, so will the importance one gives to eating roast duck or listening to Coltrane. No structural difference between symbolic utilities and others has yet been pointed out.

Nozick also cites "another reason why symbolic utilities must be treated as a separate component of a theory of decision" (34):

"[it] is not the case that a half or a one-tenth chance of realizing a certain goal always itself has half or one-tenth the symbolic utility of the goal itself—it need not symbolize that goal, even partially. ...[S]uch symbolic utilities do not obey an expected value formula" (34).

What does Nozick's observation tell us about symbolic value? If the utility of symbolizing a situation literally derived from the situation itself, then it might seem natural to think that one's actually obtaining the utility depended on the situation coming to pass. In that case, some sort of expected value formula would seem to apply: a 50% chance of realizing a certain symbolized situation would have half the value of a 100% chance. But this does not seem to hold in general.

This suggests to me that although the symbolic utility may in some metaphorical sense flow from the situation symbolized, this is best seen as no more than metaphor. The utility comes from the symbolizing of the situation, and symbolizing a situation may have value—not expected value—entirely independent of whether the symbolized situation comes to pass. Whether a given act symbolizes a certain situation may, in some cases, depend on the fact that the act is expected, more or less strongly, to bring the situation about. But when one sees the symbolizing itself as having value, one does not expect that value to be proportional to the probability that the symbolized situation will come about.5

The main question, however, is this: does the fact that symbolic value is not proportional to the probability of the symbolized situation show that it cannot be fit into standard decision theory? I do not see how it would. If the symbolic value of an action does not typically depend on the probability of the symbolized situation coming about, then it might typically be the case that symbolic utility does not vary amongst the act's possible outcomes—it is realized whichever outcome occurs. But so far, we have seen no reason to think that there is any problem with taking all of a certain act's possible outcomes to include the same symbolic utility.6

The final argument I found for treating symbolic value differently from all other value occurs in Nozick's discussion of Prisoner's Dilemma cases. It begins as follows:

It might be thought that if an action does have symbolic utility, then this will show itself completely in the utility entries in the matrix for that action (for example, perhaps each of the entries gets raised by a certain fixed amount that stands for the act's symbolic utility), so that there need not be any separate SU factor. Yet the symbolic value of an act is not determined solely by that act. The act's meaning can depend on what other acts are available with what payoffs and what acts are available to the other party or parties. What the act symbolizes is something it symbolizes when done in that particular situation, in preference to those particular alternatives. (55)

So far, this seems like a correct observation about symbolic value. But the question to be asked is why this warrants giving symbolic value special treatment. Nozick writes that it
shows an act's symbolic utility "is not a function of those features captured by treating an act in isolation, simply as a mapping of states onto consequences" (55). But he appends a note acknowledging that "an act cannot be reduced this way, even apart from issues involving its possible symbolic value" (190, fn. 17), so this cannot be his argument for treating symbolic utility separately.

The point that the symbolic aspects of an act may depend in a general way on context does not distinguish symbolic aspects from countless others. Whether a particular gun firing is an instance of target practice, hunting, or homicide depends in part on what the gun is pointed at. The lesson such examples hold for decision theory is that acts must be individuated finely enough to capture such distinctions.

Now the contextual elements that help determine symbolic value include, as Nozick points out, the range of alternative acts open to the agent. This particular sort of contextual dependence might be thought to present a special problem, by threatening the decision theoretic principle that preference between two acts should not depend on whether some third act is available. But three points should be made about this worry.

First, even this type of contextual dependence is not peculiar to symbolic value. Whether an action is rude or polite, courageous or cowardly, moral or immoral often depends on the range of options the agent has open to her. So if this type of dependence poses a difficulty for decision theory, it is a general one, and does not call for treating symbolic values differently from the rest. Second, it does not seem that this sort of difficulty would be solved even by taking all alternative-act-dependent sources of value and including them under a separate term in the DV equation. The relevant decision-theoretic principle would still be violated. Finally, and most importantly, there is no apparent barrier to simply applying the standard decision-theoretic treatment described above to this special case, by individuating acts finely enough to capture their symbolic aspects. We may readily distinguish between, say, a refusal to confess that symbolizes cooperation and a refusal to confess that has no such symbolic significance. Thus the fact that symbolic value depends on the range of options available to the agent does not seem to pose a new problem for standard decision theory.

Nozick goes on to describe the problem in a way that sounds more formal:

An act's symbolic value may depend on the whole decision or game matrix. It is not appropriately represented by some addition to or subtraction from the utilities of consequences within the matrix. (55)

It is not clear exactly what problem Nozick has in mind here. Suppose that the suggestion is that the symbolic utilities depend on the array of choices and non-symbolic utilities that a decision matrix might represent. This is perhaps the most natural interpretation of Nozick's claim about what symbolic utility depends on. But there is no obvious barrier to taking such a matrix of choices and non-symbolic utilities and simply adding the appropriate symbolic utility increments to the non-symbolic utilities already represented.

Perhaps, then, the suggestion is that the symbolic utility at some particular point in the matrix depends on the symbolic utilities of other points, which in turn depend on the symbolic utility of the original point. This possibility is not explicitly raised by Nozick, but would seem to present a clear prima facie difficulty. However, the difficulty here seems not to be caused by attempting to treat symbolic utilities along with others. It seems to lie in a particular conception of how symbolic utility comes about. It is a metaphysical
problem, and there is no reason to think that it could be solved or avoided by treating symbolic utility as a separate quantity. Moreover, if one succeeded in defining a separate symbolic utility quantity, it is hard to see why this shouldn't simply be added as an increment to the appropriate non-symbolic utility entries in the matrix.

Nozick goes on to offer a somewhat different perspective on the problem:

...if the reasons for doing an act A affect its utility, then attempting to build this utility of A into its consequences will thereby alter the act and change the reasons for doing it; but the utility of that altered action will depend on the reasons for doing it, and attempting to build this into the consequences will alter the reasons for doing that now doubly altered act, and so forth. Moreover, the utilities of an outcome can change if the action is done for certain reasons. What we want the utilities of the outcomes to represent, therefore, is the conditional utilities of the outcomes given that the action is done for certain reasons. (55)

There may be an interesting difficulty here for decision theory (though its implications are not traced out in The Nature of Rationality). But does it present an argument for treating symbolic values separately from all the rest? It might (though that would require argument), if symbolic value of an act were the only kind of value which depended on the reasons for which the act was performed. But it isn't. The moral values of actions typically depend on the reasons for which the actions are done. And the same goes for the everyday kind of utilities often considered in decision theory. If one goes to a concert to be seen, or drinks a fine wine to get drunk, or has sex to earn money, one is likely to miss out on the aesthetic, gustatory, or erotic value that those acts could provide when done for other reasons.

Thus Nozick's argument does not seem to point to any problem peculiar to including symbolic values in standard versions of decision theory. And to the extent that he has raised difficulties for the treatment of a certain wider class of utilities, he has not made clear how removing them from the standard utility formulations would solve that problem. There may, of course, be more to this last line of argument than is evident in Nozick's brief discussion. Or there may be some other reasons, not yet articulated, for thinking that symbolic value must be given special treatment. But to my mind, no substantial case for such radical reform has been made in The Nature of Rationality.

One last question about symbolic utility remains. Recall that Nozick wants to use symbolic utility to rationalize the cooperative response in Prisoner's Dilemma-like situations. Would we lose this possibility if we failed to treat symbolic utility as a separate quantity in decision theory? There is no obvious reason for thinking that we would. After all, if a person values symbolizing cooperation, and this is taken as part of her overall utility, she may still cooperate rationally for this reason. Thus no reform of decision theory is needed to allow symbolic considerations to give reasons for cooperative responses in Prisoner's Dilemma-like situations. A parallel point can be made about non-symbolic straightforward ethical motivation. It is clear that there are situations which, when looked at from the point of view of purely self-interested motivation, constitute Prisoner's Dilemmas, yet in which a person whose goals transcend self-interest may rationally cooperate.

Of course, once we represent utility fully, in whichever way we do this, there will remain situations in which each player pursues the strategy which maximizes her utility (where this is understood as not being limited to self-interest), and this still leads to both
players achieving less utility than they would have if both had acted in a non-utility-maximizing fashion. True, adding an agent’s symbolic values into the equation, like adding her moral values, may decrease the number of cases exhibiting Prisoner’s Dilemma structure. But if the existence of Prisoner’s Dilemma situations is seen as paradoxical, the paradox persists.

4. Rational Belief and Credibility

The above discussion has concentrated on practical rationality. But The Nature of Rationality includes as well a theory of the other main topic generally included under its title: rational belief. Of course, if beliefs are considered as actions, one could presumably use decision theory to determine which beliefs it would be most rational, in the practical sense, to adopt. But, as in the case of Pascal’s wager, a belief that is rational in this sense to adopt might well fail to be a rational belief in the epistemic sense that forms much of the subject of epistemology. Nozick distinguishes between these senses of rationality by reference to the example of a mother presented with evidence that her son has committed a crime—evidence that makes his guilt more credible than his innocence. Believing him innocent might be the rational thing for her to do, since the thought of his guilt might be extremely painful for her. But the proposition that he is innocent would not be a rational one for her to believe. It is this second sort of rationality that Nozick seeks to illuminate.

Nozick’s account of rational belief rests on two pillars. The first assesses the credibility of the various candidates for belief. A statement’s credibility value for hypothesis h is to be seen as “an ideal assessment that duly weights all of the reasons for and against h” (84). The assessment—which Nozick only sketches, and sees as evolving and being adjusted as time goes on—is to incorporate several traditional measures of epistemic appraisal: Bayesian probabilities, Popperian methodological maxims, and a way of assessing whether reasons for a given h are undercut. It also incorporates a novel measure of explanatory support devised by Nozick. The measure is modeled on Bayes’ theorem, but the traditional conditional probabilities are replaced by a notion symbolized by ‘→’ in the scheme below:

\[
\text{Measure } (h/e) = \frac{\prod_{i=1}^{n} \text{prob}(h_i \rightarrow e) \times \text{prob}(h_i)}{\text{prob}(h \rightarrow e) \times \text{prob}(h) + \text{prob}(C)} \]

One problem with assessing this measure is unclarity about the intended significance of the arrow notation. Nozick sometimes gives a simple subjunctive interpretation, explaining \(\text{prob}(h \rightarrow e)\) as “the probability that if h were true, e would be true” (82). But at other times he gives an explicitly causal interpretation, explaining \(\text{prob}(h_1 \rightarrow e)\) as “the probability that if \(h_1\) were true it would give rise to e” (195, fn.29). Although Nozick seems to use these notions interchangeably, they are not equivalent. The probability that if I were to feel stuffed, I would have eaten a big meal is quite high, while the probability that if I were to feel stuffed, it would give rise to my eating (or having eaten) a big meal is not.

Some evidence for the simple subjunctive interpretation is provided by Nozick’s wondering whether his measure might measure “the probability that if e were true \(h_1\) would be true, \(\text{prob}(e \rightarrow h_1)\)” (83). Aside from the repetition of the simple subjunctive language,
this suggestion would not make sense on the causal reading. A high probability that if h₁ were true it would give rise to e would not be expected to correlate with a high probability that if e were true it would give rise to h₁.

Nevertheless, causal formulations predominate in Nozick’s discussion. Also, the motivation for including the last term in the denominator seems to flow from the causal interpretation (on the simple subjunctive interpretation, the probability of e arising from a chance process would be included in the prob(h₁→e)). Most importantly, insofar as the measure is intended as capturing the extent to which h₁ would explain e, subjunctive correlations based on e causing h₁ (rather than h₁ causing e) would seem to be irrelevant. Thus it seems best to interpret Nozick’s proposal in accordance with his causal formulations.

The measure is intended to feed into the credibility assessments of hypotheses, and this may seem reasonable inasmuch as hypotheses are standardly taken to gain credibility when they explain our evidence. Of course, we would not want to give credibility to h₁ in virtue of its explaining e unless we had substantial credence in e itself. Nozick’s references to e as “data” or “evidence” indicate that he intends e to be taken as accepted. One might naturally want to weaken this a bit, by dropping the requirement that e be accepted, and instead multiplying the above measure by prob(e) when feeding it into the credibility measure. But whichever way one treats the measure, it is clear that h₁ should get credit for explaining e only to the extent that we are confident of e itself.

This point, however, brings to the fore a serious difficulty. Suppose first that e itself is not yet believed, but that we believe that h₁ would be a likely cause of e, if e were true. (For example, e might be “the barometer is falling”, and h₁ “the atmospheric pressure is decreasing,” and we might not have checked the barometer yet.) Now if we believe that h₁ would be a good explanation (in this sense) of e, it would seem that this will typically have a big impact on their probabilistic interrelations. In the example, our beliefs about the likely causal relations between the atmospheric pressure and barometer readings will ordinarily be reflected in the fact that our conditional probability for h₁ given e is much higher than our unconditional probability for h₁.

Now in cases like this, according to the standard Bayesian mechanisms, it will typically be the case that our coming to believe e will serve to raise the probability of h₁. (Nozick himself notes that “the previous explanatory successes of an hypothesis will…affect its prior probability as it enters into this formula” (83).) Suppose, then, that e is strongly believed already, as Nozick’s measure must require. In that case, our belief in e will already have had its effect in raising h₁’s probability.

This is problematic in two ways. The first derives from the fact that the current probability of h₁—which already reflects its explaining e—is one of the inputs to Nozick’s measure. The measure is supposed to determine the extent to which h₁ is supported by explaining e. Of course, in assessing how good an explanation h₁ is of e, one must take account of how likely, on other grounds, h₁ is: if h₁ is in itself highly improbable, it isn’t a very good explanation of e—even if, were h₁ true, it would likely give rise to e. (This is Nozick’s motivation for making prob(h₁) an input to the measure.) But it would seem that this assessment of the probability of h₁, for the purpose of evaluating how much support it receives by explaining e, cannot legitimately depend on our already judging h₁ more probable because it explains e. But it seems that once we have a strong belief in e, the current probability of h₁ gets adjusted in a way that reflects this very judgment.⁸

The second problem is not a difficulty with the measure per se, but with its use as a
factor in Nozick’s multi-factor credibility assessment. Recall that this assessment includes the current probability of \( h \) as a separate factor. But given that the current probability of \( h \) already reflects its explaining \( e \), it would seem like double counting to increase its credibility further by using Nozick’s measure. Even if the first problem with the measure could be avoided, then, it is not clear that the measure would have a legitimate function in Nozick’s credibility assessment.

Let us, however, put this difficulty aside, and see how the credibility assessment functions in Nozick’s account of rational belief. The second pillar supporting Nozick’s account we have already encountered: the notion of Decision Value. Nozick’s basic account can be expressed in three conditions.

Believe \( h \) if and only if:

1. no statement incompatible with \( h \) has a higher credibility value (CV) than \( h \) does; and
2. the CV of \( h \) is high enough, given the kind of statement \( h \) is; and
3. the decision-value (DV) of believing \( h \) is at least as great as the DV of having no belief about \( h \).

Nozick applies this account to Kyburg’s “lottery paradox” example, in which an agent considers a very large lottery with one winning ticket (see Kyburg (1970)). Nozick’s account follows Kyburg’s approach of allowing a rational agent to believe of each ticket that it will not win, but also to believe that one of the tickets will win—a clearly inconsistent set of beliefs. But, like Kyburg, Nozick wants to rule out rational beliefs that are pairwise inconsistent. (This follows from clause 1 of the above account, along with a special provision excluding belief in incompatible propositions with equal credibility values.)

Nozick adds one more principle to his account, to allow formulation of new beliefs on the basis of logical inference from the beliefs sanctioned by the basic account above. Having allowed inconsistent belief sets already, he cannot allow unrestricted use of logical inference on beliefs without sanctioning belief in everything. Thus he rounds out his account with the following principle:

Believe \( q \) because it is inferred from premises \( p_1, \ldots, p_n \) in an explicit deductive inference only if each of these premises \( p_i \) is believed and only if their conjunction \( p_1 \& p_2 \& \ldots \& p_n \) also is believed. (92)

Perhaps the most distinctive feature of Nozick’s system is the mixture of pragmatic with purely epistemic considerations in the basic account. A belief is rational just in case it passes the epistemic tests in the first two conditions and the Decision Value utility of believing it is at least as high as that of withholding belief. This is particularly interesting in view of Nozick’s having taken care to distinguish the notion of “rational belief in \( P \)” from the pragmatic notion of “believing \( P \) being the rational thing to do”. Why should we think that the pragmatic factors involved in DV are relevant to rational belief?

It might be that the pragmatic dimension helps explain our intuitions about certain cases of belief. In the case of the mother confronted with evidence of her son’s guilt, Nozick says that it would be irrational for her to believe her son innocent, because that proposition has a lower credibility value than the incompatible proposition that he is
guilty. But it would not be irrational, says Nozick, for her to withhold belief on the matter, given that believing her son guilty would have lower DV utility for her than withholding. This verdict is consonant with clauses 1 and 3 of Nozick’s account.

But how much support can Nozick’s account take from this intuition? It may be that we are less inclined to cry “irrational!” when told about the mother’s withholding belief. But when one keeps in mind the difficulty of keeping “P is rational to believe” and “believing P is the rational thing to do” clearly distinguished, along with the fact that withholding belief in this case is less irrational (in a purely epistemic sense) than believing the son innocent, it seems that not too much weight can be put on the intuition Nozick describes.10

Furthermore, the full implications of Nozick’s account are far less intuitively palatable. Suppose that the mother in our example decided to believe, in accordance with the evidence, that her son was guilty, despite the fact that having this belief was so painful that withholding would have had greater overall utility for her. Is her belief irrational? To my mind, the suggestion that the mother be convicted of epistemic irrationality here is absurd. “Wishful withholding” may not strike us as being so clearly irrational as wishful thinking. But no account of epistemic rationality should fault an agent who lets the evidence guide her beliefs despite the personal costs of doing so. Indeed, I would think that such agents furnish us with paradigms of epistemic virtue.11

One could, of course, tinker with Nozick’s account to avoid this sort of example. For example, one might require belief in accordance with the evidence when such belief has greater DV-utility (as Nozick does), but then allow either belief or withholding when the DV-utility stacks up the other way. But before doing this, one might ask how plausible it is that the practical value of holding a belief helps determine its rationality.

Consider our intuitions about the mother who believes in accordance with the evidence, despite the pain this causes her. Our intuitions that her belief is rational have their source in our conception of epistemic rationality—a conception that at least seems to involve divorcing the pursuit of truth from other concerns. But though Nozick’s account of rational belief does give pursuit of truth a central place, it also gives a pivotal role to practical considerations that are blatantly non-truth-related.

In Nozick’s discussion of cognitive goals, he does point out the difficulty of saying just what our cognitive goals are (believing any old truths? avoiding error? achieving explanatory power?). He also suggests that our concern with truth is itself ultimately (evolutionarily) grounded in the fact that true beliefs are instrumental to our other goals. But this does not, it seems to me, do much to support the claim that our notion of “p being rational to believe”—once we distinguish this notion from “believing p being the rational thing to do”—answers to the sort of distinctively non-truth-related, and, indeed, non-cognitive, goals that figure in DV.

In the end, then, Nozick says very little to support the claim that purely practical considerations play an important role in determining what it is rational for us to believe. There is, undoubtedly, a lot more to say on this issue.12 But those who are initially suspicious of pragmatic accounts of epistemic rationality are unlikely to have their suspicions allayed by this book.

5. Conclusion

There are a great many ideas put forward in The Nature of Rationality that I have not touched on. Examples from the sections of the book I’ve concentrated on include a discussion of the
roles played in our lives by principles; a defense of the propriety of taking sunk costs into account in our practical calculations; and advocacy of the study of artificial intelligence and cognitive science to give us insight into rational belief formation.

The book also includes a chapter on the evolution of rationality. Nozick wants to bring together two strands of thought about reasons: that they are grounded in a priori relations, and that they are grounded in contingent factual relations. He suggests that reasons are grounded in factual relations that have, through evolution, come to seem to be self-evidently supportive. The chapter includes an analysis of evolutionary fitness and one of biological function, and a speculative explanation of the intractability of certain philosophical problems (our rational faculties evolved to take certain assumptions for granted, and are thus incapable of justifying them).

A final chapter argues that not all rationality is instrumental. After noting that some rationality conditions on preferences (e.g. transitivity) are widely accepted, Nozick goes on to put forward 22 more. For example:

V. The person prefers that each of the preconditions (means) for her making any preferential choices be satisfied, in the absence of any particular reason for not preferring this. (142)

Nozick makes no serious attempt to defend any of these particular principles (indeed, all 22 are put forward in 10 pages). Rather, he intends them to illustrate that there is ample room for conditions of this sort in the theory of rationality. Also included here is a section on philosophical heuristics (16 in less than five pages); a short example:

“8. Examine extreme cases, consider what will result if some parameters are set at zero or at infinite value, and then reconsider your intermediate case in the light of this extremal behavior” (170).

The chapter also contains some reflections on the role of imagination in rationality.

Clearly, this is a book containing more than its share of positive philosophical suggestions. And given its author’s imagination and intelligence, it is no surprise that many of these suggestions are intriguing. I suspect that few readers will finish this book without thinking over some interesting ideas in new ways. Moreover, the book’s lack of wrangling over details makes it eminently readable.

The book’s wealth of ideas and non-defensive style help give it, as noted above, some of the feel, and the benefits, of philosophical conversation. But those benefits have not been achieved without cost. The chief drawback of many philosophical conversations is that the ideas go by too fast. Without careful and detailed—even tedious—consideration of arguments pro and con, it is hard to distinguish the truly valuable ideas from those which merely present an appealing face. This drawback also constitutes, to my mind, the main weakness in The Nature of Rationality. Though it clearly contains more in the way of argumentation than would any casual philosophical conversation, still, its ideas often pass by quite quickly.

How troubling one finds this will depend on one’s concerns. Many readers will be grateful for Nozick’s lively pace. Others, though, will feel frustrated. The Nature of Rationality is a provocative book. But many readers will wish that it would slow down a bit, to allow its suggestions to be explored and defended in greater detail.
Notes

1 I would like to thank Stephen Jacobson, Mark Kaplan, Hilary Kornblith, Arthur Kuflik, Don Loeb, Derk Pereboom and Robert Nozick for helpful discussions and/or comments on earlier drafts.

2 Nozick does not specify what weights should be given to the two components of decision-value, though he asserts that “each must be given its respective due. The weights, then, are not measures of uncertainty but measures of the legitimate force of each principle” (45). Nozick suggests that the right weighting will depend on the character of the world we inhabit. The relationship between the character of our world and the EEU/CEU controversy is discussed in more detail below.

3 Perhaps I have dogmatic colleagues. Certainly these people are professional philosophers who have reached a certain conclusion in the Newcomb case. But I see no reason to think that the intuitions of philosophically unsophisticated people would be more revealing.

4 Nozick speaks simply of “Prisoner’s Dilemma situations.” My scruple here is due to the fact that if symbolic value gives the cooperative action greater utility for the prisoner, it is not obvious that the real Prisoner’s Dilemma structure persists.

I should also note that part of Nozick’s approach to Prisoner’s Dilemmas involves his endorsement of an EEU component in decision theory. When one prisoner believes that the other prisoner is relevantly very similar to him, then his choosing the cooperative strategy will serve as evidence of the other prisoner cooperating. Since each prisoner does better when the other cooperates, cooperating may have greater EEU. Nozick’s discussion of this matter largely parallels his treatment of the Newcomb problem discussed above.

5 I would thus reject part of Nozick’s description of the relation between the utility of the symbolized situation and the symbolic utility that flows from it. At one point he writes:

Symbolic utility is not a different kind of utility, standing to standard utility in something like the way that metaphorical meaning stands to literal. Rather, symbolic utility is a different kind of connection—symbolic—to the familiar kind of utility. It stands alongside the already familiar connections, the causal and the evidential. (48)

It does seem right to say that symbolic value doesn’t stand to all other value as metaphorical meaning stands to literal meaning. But it is misleading, I think, to take symbolic value as merely a “connection” to non-symbolic value. The fact that the symbolic value of an act does not depend on even the possible occurrence of the symbolized situation suggests strongly that symbolizing is best seen as a source of value for us, not just a connection to some other value. I thus see no reason to assume, as Nozick does at one point, that symbolic utility is never greater than the non-symbolic utility of that which is symbolized.

6 Nozick sometimes talks of the symbolic utility arising when the outcome of an act, rather than the act itself, symbolizes some further situation. This might work in two ways. The first would be that the symbolic utility of a possible outcome of an act would be realized by the act even if that outcome didn’t actually occur. (The strength might depend on the probability that the outcome would have occurred, but it might not, as Nozick points out.) This kind of utility would behave like utility attached to the act itself, and might be made part of the act’s CEU (or EEU) directly, as described in the text.

But it might be thought that the symbolic utility could be attached more intimately to the outcome itself, so that it would be realized only if that outcome occurred. In that case, it would seem that something like an expected symbolic utility measure might be needed to guide our decision after all. Perhaps, given Nozick’s framework, we’d need CESU and EESU. But again, why not simply add the symbolic utility to the rest of the utility of the outcome, and having it feed into the act’s CEU (or EEU) in the ordinary way?

7 This last quantity, prob(ce), is intended to represent “the chance hypothesis about e...the denial that there is any h, such that h did, probabilistically, generate e” (83). No analogue of this quantity appears in Bayes’ theorem, because the hypotheses considered in the denominator are stipulated to be exhaustive. No such stipulation is made in Nozick’s formula.

8 Thus although Nozick believes that his measure escapes the Bayesians’ “problem of known evidence” (195, fn29), a close relative of the problem can be seen lurking just beneath the measure’s surface.

9 This is a curious formulation. The requirement that p1 - p0 be believed individually would appear
to be vacuous, given that belief in their conjunction is required. (Even if one began by believing the conjunction but not the conjuncts, one would be entitled by the rule itself to believe the conjuncts.) If we strip away the vacuous clause, we see that the rule essentially permits belief via logical inference from single premises only. This brings out how Nozick’s rule is remarkably similar—though it is not identical—to the “Weak Deduction Principle” at the heart of Kyburg’s system: “If S is a body of reasonably accepted statements, and s₁ belongs to S, and s₁ ⊻ s₂ is a theorem of our underlying logic, then s₂ belongs to S” (op. cit. 55).

Moreover, in cases of extremely compelling evidence, as Nozick seems to acknowledge (87, fn), withholding may well seem intuitively irrational.

It would of course be inhuman to insist that people in the mother’s situation should always believe the very painful truths. But that is just to say that it would be inhuman to insist that epistemic rationality trump all other human considerations.

For an extended defense of an account of rational belief in which truth per se plays no role, see Stich (1990). This aspect of Stich’s account is discussed critically in Goldman (1986). Kornblith (1993), and Jacobson (1992).

References


ANDREW NEWMAN
University of Nebraska at Omaha

The first few chapters of this book draw some general conclusions about a non-Humean theory of causality and scientific explanation from complex examples taken from econometrics and physics; the later chapters describe a general metaphysics of causality for science. The book ends with an interesting application to the Einstein Rosen Podolsky paradox.

Despite the fact that the book is about causal powers, Nancy Cartwright keeps her