

DISPOSITION-BASED DECISION THEORY

Justin C. Fisher
 Southern Methodist University
<http://www.justin-fisher.com/>

ABSTRACT. I develop and defend a version of what I call Disposition-Based Decision Theory (or DBDT). I point out important problems in David Gauthier's (1985, 1986) formulation of DBDT, and carefully develop a more defensible formulation. I then compare my version of DBDT to the currently most widely accepted decision theory, Causal Decision Theory (CDT). Traditional intuition-based arguments fail to give us any strong reason to prefer either theory over the other, but I propose an alternative strategy for resolving this debate. I argue that we should embrace DBDT because it does better than CDT at the work that we, as a matter of empirical fact, commonly call upon a notion of rationality to do.

1. INTRODUCTION

A *decision theory* (or rational choice theory) specifies which choices it is permissible for an agent to make, given her beliefs about her current situation and given her desires regarding potential outcomes in that situation. In other words, a decision theory describes a function from what I will call *scenario descriptions* to sets of choices, and a *good* decision theory will map each scenario description to the set of choices that would be *rationaly permissible* for any agent whose current beliefs and desires are reflected in that scenario description.

Of course there are extremely many functions that map scenario descriptions to sets of choices, and almost all of these

candidate functions would constitute very bad decision theories. As we will see, traditional lines of argument based on formal and intuitive considerations cannot help us to choose between several importantly different candidates. My first goal in this paper will be to describe a new argumentative strategy that may give us good reason to embrace a particular candidate, even in cases where more traditional strategies fail. My new strategy calls upon us first to seek an empirical understanding of the important work that people actually use 'rationality'-talk for, and then to favor whichever candidate is best capable of doing this work.

My second goal will be to develop and defend a version of what I take to be the best candidate, what I call Disposition-Based Decision Theory (or DBDT). I will begin by considering David Gauthier's (1985, 1986) disposition-based theory of rational choice in multi-player games. I point out important problems in Gauthier's formulation, and then carefully develop a more defensible version of DBDT.

My final goal will be to compare my version of DBDT to the candidate that is currently most widely accepted, Causal Decision Theory (or CDT). I will argue that DBDT and CDT are formally on a par, and that intuitive considerations also cannot give us good reason to embrace one of these theories rather than the other. However, I will argue, my proposed argumentative strategy *does* distinguish between these: we should embrace DBDT because it does better than CDT at doing the work that we, as a matter of empirical fact, commonly call upon a concept of rationality to do.

2. ARGUMENTATIVE STRATEGY

Two observations may help motivate my proposed argumentative strategy. First, there are empirical facts about what benefits are delivered by our use of a concept of rational

decision-making. And second, we have reason to embrace whichever candidate theory provides a notion that best sustains such beneficial usage. Given the choice between a notion that would do well at delivering the sorts of benefits we call upon our concept of rationality to deliver and one that would do poorly, we have clear reason to favor the one that does well.¹

Hence, this strategy calls upon us to develop arguments of the following general form:

- P1. We use our rationality-talk for job J.
- P2. Of the various precise notions that we might plausibly associate with the term ‘rational’, notion N does best at performing job J.
-
- C. Therefore, we should accept that the term ‘rational’ means notion N.

Paradigm cases of rationality-talk involve people’s using the term ‘rational’ in ways that clearly have to do with prudential, means-end reasoning. Such paradigm cases might take one of the following forms:

- (a) “It was rational of agent A to choose C in scenario S.”
- (b) “Agent A is (generally) rational.”

Of course, such paradigm cases are quite rare in common parlance – much less frequent than our employment of

concepts that might usefully be explicated by a good theory of rational choice. So, we should also count many other ordinary utterances (when made in ways that clearly have to do with prudential means-end reasoning) as instances of rationality-talk. Some examples:

- (c) “It was clever/smart/right/prudent/good of agent A to choose C in scenario S.”
- (d) “Agent A is clever/smart/wise/prudent.”
- (e) “Agent A is good/smart at making decisions.”
- (f) “Agent A always knows what (or knows the right/best thing) to do.”

This list can serve as a first approximation – subject to extension and revision on empirical grounds – of the sorts of occasions where we employ a concept of rationality that might benefit from a unified explication.

One might read the conclusion of my argument in two ways. On a strong reading, it offers an analysis of what the term ‘rational’ *meant all along*. I am quite sympathetic to this strong reading, though I grant that it rests upon a controversial “meaning is use” semantic presumption. Rather than attempting to defend such a presumption here, I would like to note that a more plausible “meaning ought to follow use” reading of the conclusion will suffice for present purposes. This reading allows that the term ‘rational’ may *thus far* have been quite ambiguous or even misguided in its meaning, but concludes that, so long as we intend to continue using this term for the purposes we’ve traditionally used it for (i.e., for job J), we have good reason *henceforth* to define it to mean the disposition-based notion of rationality that I propose below. This reading of the conclusion, by itself, would have striking

¹ In other work (Fisher, in prep), I call this general strategy *Pragmatic Conceptual Analysis*, and discuss its relative merits *vis a vis* other philosophical methodologies.

implications for rational choice theory, game theory, economics, and moral and political philosophy.

Let us now turn to the notion of *job* employed in my argument. As a first approximation to this intuitive notion, we may say that a group of people uses practice P to do job J if their use of P often enough accomplishes J and the benefits produced by accomplishing J are what makes it worth engaging in P as they do.

For example, medieval people's use of cooking fires often enough accomplished the job of killing harmful bacteria in their food, and it was these beneficial effects that made these practices worth doing. Notice that cooking fires were used for this job long before people knew about the existence of bacteria. The fact that this was a job that cooking fires were being used for does not depend in any direct way upon these people's *intuitions* nor upon their *explicit intentions* regarding this practice.

Similarly, I take it to be an empirical fact that we commonly use rationality-talk for certain jobs and not for others, and I expect that this empirical fact does not depend in any direct way upon our intuitions or even upon our explicit intentions regarding our rationality-talk. To determine what jobs our rationality-talk is being used for, we must attend carefully to the ways that this talk is used and the important benefits which we (often enough) thereby accrue.

I will now offer two plausible hypotheses regarding the jobs for which we use our rationality-talk. First, we use our rationality-talk to designate choices and ways of choosing that are worth aspiring towards. We tag choices and ways of choosing as 'rational' in order to make ourselves and the people we care about more likely to make future choices in a similar way. Conversely, we tag choices and ways of choosing as 'irrational' in order to make ourselves and our hearers *less*

likely to make future choices in the ways in question. In both cases, the job being fulfilled (often enough) by this practice is the job of shaping our choice-making dispositions so that we will be more likely to satisfy our preferences in the scenarios in which those dispositions may become manifest. These sorts of 'rationality'-talk are worth doing primarily because they tend to perform these jobs.

A second plausible hypothesis is that we commonly use rationality-talk to assess an agent's decision-making capabilities, so that we might know whether to ask her for advice, or know whether to invest our money with her, or know how successful to expect her ventures to be. Such assessments are worth doing because, often enough, they are correct and thereby yield important dividends. Hence, on the intuitive notion of *jobs* presented above, making such assessments is quite probably one of the jobs we use our rationality-talk for.

In slogan form, we may summarize these hypotheses by saying that we use our rationality-talk (1) to help ourselves, our friends, and especially our children to become more successful decision-makers, and (2) to indicate which people we may expect to make or suggest successful decisions. I think most 'work' done by our rationality-talk may be subsumed under one or the other of these job descriptions. This is, of course, an empirical claim that could conceivably be refuted by empirical findings, a possibility that I will return to below. In the meantime, if we find that one candidate decision theory does better than its rivals at the above jobs, then we will have strong *prima facie* reason to embrace that candidate.

3. TWO DECISION SCENARIOS AND TWO JOB CANDIDATES

My next task will be to present what I think are the two strongest candidates for the jobs described above. In presenting

these candidates, it will be helpful to have in mind a case regarding which the two candidates disagree:

Newcomb's Problem.² You must choose either (one-box) to take just an opaque box or else (two-boxes) to take *both* that opaque box *and* a transparent box that clearly contains one thousand dollars. You can't see what is in the opaque box, but you know its contents were determined as follows. A few minutes ago, you and your surroundings were carefully examined by a very good predictor of agents' behaviors. If this inspection made the predictor confident that you would choose one-box, then she put one million dollars in the opaque box; otherwise she put nothing in it. Assuming that your only relevant desire is to accumulate dollars, what would it be rational for you to choose?

Newcomb's problem is a reasonable idealization of actual cases where predictors of human behavior (like romantic partners or hiring committees) determine whether or not an agent will receive a highly desirable payoff on the basis of their predictions of that agent's future choices. So long as the difference in potential payoffs is quite large, the predictors needn't be much better than chance to make these actual cases puzzling in the same way that Newcomb's problem is. Still, to ensure we aren't just dwelling upon bizarre cases with little practical import, it will be helpful to have in mind a second case as well:

The Psychologically-Similar Prisoners' Dilemma. You must choose either to cooperate or to defect. Another prisoner in a distant cell is faced with a choice just like yours. *This other prisoner was recently*

selected to play this role because of the psychological similarity between you – a psychological similarity which makes it very likely that the other prisoner will make the same choice that you make. If both of you cooperate you each will be sentenced to 1 year in prison. If both of you defect you each will be sentenced to 3 years in prison. If one cooperates while the other defects, then the cooperator must serve 5 years in prison, while the defector gets off scot-free. Assuming that your only relevant desire is to minimize your years in prison, what would it be rational for you to choose?

Despite their somewhat artificial structure, prisoners' dilemmas are commonly taken to be a reasonable starting point for understanding many dilemmas of human cooperation. What makes the *psychologically-similar* prisoners' dilemma especially puzzling is the stipulation of psychological similarity italicized above. For actual human agents, this stipulation is at least approximated by our societies' mechanisms of social assortment which (plausibly) produce at least a weak tendency for like-minded individuals to be paired in cooperative dilemmas together. Hence, this case is related to many cases that real people often face.

We may use these two decision scenarios to contrastively illustrate (what I take to be) the two strongest candidate theories of rational decision-making. The first candidate gives a central role to the following heuristic.

Causal Dominance Heuristic: Usually, one rationally ought to make a given choice if, regardless of how things outside one's control turn out to be, one would be better off if one made that choice than one would be if one made the other choice.

² The *locus classicus* for Newcomb's problem is Nozick (1969).

In Newcomb's problem, regardless of how much money the predictor has hidden in the opaque box, choosing two-boxes will deliver \$1000 more than would choosing one-box in the very same circumstance. Hence, this heuristic suggests, the rational choice is two-boxes. Similar reasoning concludes that in any prisoners' dilemma you should defect; for, regardless of what the other prisoner chooses, you will be better off defecting than you would be cooperating in the very same circumstance.

The causal dominance heuristic is endorsed by our first, and currently most popular, job candidate, *Causal Decision Theory* (hereafter CDT). According to CDT, it is rational for an agent to choose whichever option she believes will *cause* the best expected change in her utility, with factors outside her causal control considered as fixed constraints.³

Despite the intuitive attractiveness of the causal dominance heuristic, many people also have persistent intuitions to the contrary. For example, in the above decision scenarios, one-boxers and cooperators would consistently and predictably receive much better payoffs than would two-boxers and defectors going into similar scenarios. Hence, there is clearly *something* to be said in favor of one-boxing and cooperating, and many people have the intuition that the term 'rational' is supposed to track this *something*, even when this forces us to diverge from the causal dominance heuristic.

The classic attempt to formalize this intuition was *Evidential Decision Theory* (EDT), which held that it is rational to choose whichever choice is correlated with the highest expected payoff. Many authors have correctly noted that EDT gives clearly wrong prescriptions in temporally

extended cases like the Smoking Cancer Problem, which I will discuss below. Rather than dwelling on EDT, let us consider what I take to be a much stronger way of formalizing one-boxer intuitions.

This stronger view employs a *disposition-based* conception of rationality; it holds that what should be directly assessed for 'rationality' is *dispositions to choose* rather than *choices* themselves. Intuitively, there is a lot to be said for the disposition to choose one-box in Newcomb's problem – people who go into Newcomb's problem with this disposition reliably come out much richer than people who instead go in with the disposition to choose two-boxes. Similarly, the disposition to cooperate in a psychologically-similar prisoners' dilemma reliably fares much better in this scenario than does the disposition to defect. A disposition-based conception of rationality holds that these intuitive observations about dispositions capture an important insight into the nature of practical rationality.

My task in the coming sections will be to carefully spell out a robust decision theory based on this disposition-based conception of rationality and to argue that it does better than CDT at the important jobs identified above.

4. PROBLEMS WITH GAUTHIER'S APPROACH

Let us begin by considering the disposition-based approach proposed by David Gauthier (1985, 1986). One might construe the present paper as an attempt to develop Gauthier's approach beyond cases of multi-player games, and to provide further support for it. But first, we must extricate the disposition-based conception of rationality from the mistaken use Gauthier attempts to put it to.

³ There are a number of different ways of formalizing this intuitive idea. See Skyrms (1982), Lewis (1981a), Joyce (1999), Pollock (2002), and my own detailed presentation in section 5 below.

Gauthier proposes a disposition-based conception of rationality in the course of defending the rationality of what he calls *constrained maximization* or CM. CM advises an agent to “cooperate if, given her estimate of whether or not her partner will choose to cooperate, her own expected utility [given that she cooperates] is greater than the utility she would expect from the non-cooperative outcome” (Gauthier, 1985, pg 81). By ‘the non-cooperative outcome’ Gauthier means the outcome one would expect if *neither player* were to cooperate. Gauthier compares CM to the strategy that always defects in cases with payoffs and utilities like those in the standard prisoners’ dilemma. Call this strategy All-D, and these cases *cooperative dilemmas*. Gauthier’s disposition-based approach suggests that we should ask: *Which of these strategies is it rational to be disposed to make choices in accordance with?* I approve of this (sort of) question, but not of Gauthier’s answer.

In a cooperative dilemma, CM advises an agent to cooperate if she thinks it likely that her opponent will cooperate as well. Clearly it is not *always* rational to follow CM. Consider the prisoners’ dilemma in which you have good reason to think your opponent will cooperate, regardless of what you do. (To make the case very stark, imagine that your opponent is a piece of cardboard with “I cooperate” printed on it.) If you cooperate in this case your expected payoff (1 year in jail) would be preferable to the non-cooperative outcome (3 years), so, according to CM, you should cooperate. But it is intuitively clear that defecting (and thereby getting off scot-free) is rational in this case, and that cooperating is *not*. Hence, it cannot be *generally* rational to follow CM.

How is it, then, that Gauthier (1985) “demonstrates” the rationality of CM? In his “demonstration,” Gauthier makes two important assumptions, neither of which is justified by his official formulation of CM. First, Gauthier assumes that we

need only consider cases in which *each* of the two agents knows the other’s dispositions. And second, Gauthier assumes that all agents under consideration will employ either CM or All-D. Given these assumptions, knowing that the other player will cooperate is tantamount to knowing that both she and you employ CM. Hence, when we restrict our consideration to just those cases that Gauthier considers, Gauthier’s claim is tantamount to the claim that it is rational to cooperate just in case you think it likely that the other player will choose much as you do.⁴ I think *this* claim is essentially correct.⁵ But demonstrating its truth clearly does not amount to demonstrating the *general* rationality of CM.

Hence, we must keep separate what Gauthier has lumped together. We will retain the general disposition-based conception of rationality and the claim that cooperation is rational in the psychologically-similar prisoners’ dilemma. But we shall not tie this to a general defense of CM.

5. ARTICULATING A DISPOSITION-BASED DECISION THEORY

I will now attempt to develop this disposition-based conception of rationality into a robust and generally applicable Disposition-Based Decision Theory (or DBDT). To facilitate understanding, I will offer some intuitive motivation for developing DBDT in the way that I do. However, my goal in this section is not to establish that one *must* develop one’s decision theory in the way I suggest – instead, it is just to lay

⁴ In his informal discussion, Gauthier says almost as much himself. “A constrained maximizer makes reasonably certain that she is among like-disposed persons before [cooperating]” (Gauthier, 1985 pg 80).

⁵ Modulo certain concerns about when and how the psychological similarity was engineered – see my resolution of the smoking-cancer problem below.

out what I take to be a good decision theory. My positive argument for accepting DBDT will be the jobs-argument introduced above, and driven home in the final section below.

My proposed version of DBDT is *formally* quite similar to CDT even though it is quite different in spirit. I will spell out these views in parallel to highlight both the formal similarities and the deep differences between them.

Since both CDT and DBDT are intended as *decision theories* each must lay out principles that specify, for any *scenario description S* – i.e., for any relevant⁶ set S of beliefs and desires that an agent might have – what choices would be rationally permissible for any agent who takes her situation to be as S describes. Each of these theories effectively suggests that this may be done by following a four-step process.

In the first step, we isolate a set of candidates to compare. CDT compares various *choices* that an agent might make.⁷ DBDT instead compares various *dispositions*, each of

⁶ For present purposes, I will limit our consideration to so-called cases of ‘choosing under risk’ – cases in which the scenario description is ‘complete’ in the sense that it is sufficient to determine the expected payoffs that various dispositions would yield in that scenario, assuming that the scenario description is accurate. Restricting our attention to situations of choosing under risk will allow us to see clearly the differences between CDT and DBDT without getting bogged down in the many interesting questions that *any* decision theory must eventually face regarding what an agent rationally should do in so-called situations of ‘choosing under uncertainty’ where she lacks a ‘complete’ scenario description.

⁷ I think the most intuitive construal of CDT is the choice-based construal employed in the main text. However, some Causal Decision Theorists (e.g., Joyce 1999) think the primary goal of decision theory is to say what choices one should *prefer to make*, rather than saying directly what one ought to choose. In these contexts, such *preferences* are commonly taken to be *dispositions* to act out particular choices. Construed this way, CDT is just as ‘disposition-based’ as DBDT. Ironically enough, we will soon see that my version of DBDT is also just as ‘causal’ as CDT. The only point of difference between CDT, on this dispositionalist construal, and my DBDT

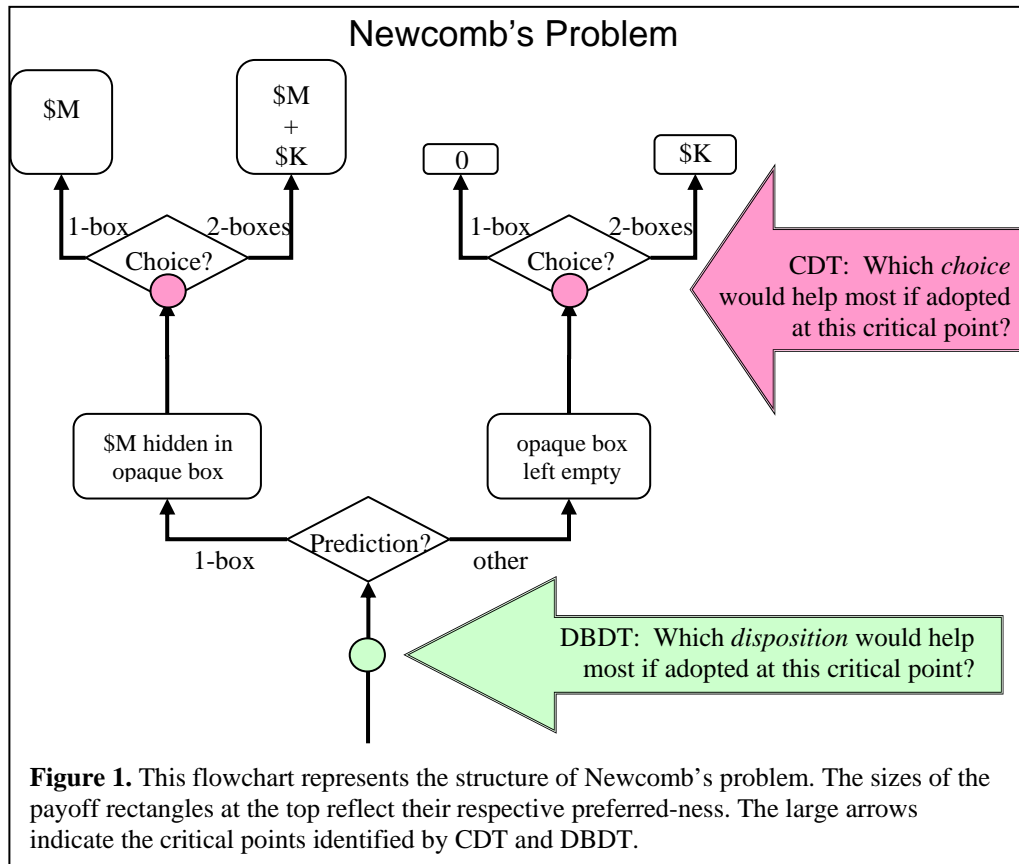
which is a disposition to choose in response to taking one’s situation to be as S describes. In simple cases, there will be one disposition for each choice.⁸

In the second step we isolate what I call a *critical point*. In later steps we will ask which candidate would do best if it were ‘adopted’ (or made) at the critical point of a scenario. For CDT the critical point is clearly *the point at which the choice will be made*, as it is the causal consequences of *making* one choice rather than another that CDT takes to be of paramount importance. (See Figure 1.)

What about DBDT? For intuitive motivation, we may return to an observation made in section 2 above. One primary use of our rationality-talk is to assess various choice-making dispositions to determine which are (or aren’t) worth adopting. This observation suggests that we may want to compare dispositions with respect to how they would fare if adopted at a point where assessments of rationality might normally shape what dispositions an agent has. Hence, DBDT defines the *critical point* of a given scenario description as the most recent time prior to the choice in question which would have been a natural opportunity for the normal shaping of dispositions. I will say more about critical points in the next section. For now, let us take it for granted that, in short-duration scenarios like Newcomb’s problem and the psychologically-similar prisoners’ dilemma, the critical point comes prior to the first

is in (step 2) what ‘critical points’ each evaluates choice-making dispositions with respect to: CDT (on the present construal) looks at what would be caused by adopting a choice-making-disposition at the time the choice is about to be made while my DBDT looks at what would be caused by adopting such a disposition earlier in the scenario. I thank Alan Hájek for pointing this out.

⁸ In more complicated cases, one might want to consider other dispositions as well – e.g., perhaps, the disposition to choose randomly among some set of options.



events mentioned in standard descriptions of these scenarios. (See Figure 1.)

In the third step CDT and DBDT each generate a new scenario description S^* just like S except that, where S describes details that wouldn't be affected by which candidate is adopted at the critical point selected in step 2, S^* instead reconstrues these factors as fixed (probabilistic) constraints on the causal processes that stem from the critical point.

For example, suppose S describes Newcomb's problem. CDT proposes that we consider the similar scenario description

S^* that brackets the details about how the contents of the opaque box were determined – for nothing the agent does now can change these – and instead says just that there is some probability p that this box now contains a million dollars, and probability $(1-p)$ that it is empty. (The value of p would be determined by the probability that the agent assigns to each of these possibilities, but, as it turns out, in Newcomb's problem it doesn't matter what value p takes.) CDT holds that whatever choice is rational in Newcomb's problem, the same choice is rational in S^* – as far as issues of rationality are concerned, S and S^* are equivalent.

Since DBDT typically identifies a critical point somewhat earlier than CDT's, DBDT usually holds that we should *not* bracket information about an agent's recent history in the way CDT does. DBDT denies that S^* , as CDT would define it, is rationally equivalent to Newcomb's problem. However, DBDT agrees that an agent's *distant* history and surroundings can be bracketed and re-construed as fixed (probabilistic) constraints without loss of rational equivalence. Hence, DBDT holds that if we instead define S^* in terms of the critical point that DBDT identifies, then S and S^* (thus defined) will be rationally equivalent.

In the fourth and final step, CDT and DBDT compare the candidates (identified in step 1) with respect to how well they would be expected to fare if adopted at the critical point (identified in step 2) of the transformed scenario S^* (constructed in step 3). Here, 'doing well' amounts to being expected to do better at achieving the desires specified in S^* . This idea might be spelled out in a number of different ways, depending, in part, upon how we imagine desires to be depicted in a scenario description.

One simple option⁹ would be to hold that a scenario description gives a preference-ordering (or partial ordering) on various ‘expected payoffs’ that a candidate might have in the scenario in question, where an *expected payoff* is a probability distribution across a set of possible outcomes. For example, consider a scenario description which says the agent may choose to bet upon the outcome of a fair coin-flip – if the flip is heads she would win \$3; otherwise she would lose \$2. The expected payoff of being disposed to reject such a bet is a certain gain of \$0; while the expected payoff of being disposed to accept involves a 50% probability of getting \$3 and a 50% probability of losing \$2. A complete scenario description would contain desires that specify which of these two possible expected payoffs is preferable. It is not difficult to imagine an agent who would find the risky chance of gaining \$3 preferable, nor is it hard to imagine an agent who would prefer not to risk her \$2. Let us assume that all relevant probabilities may be handled in a similar fashion; and hence that, for any relevant scenario description S^* , a savvy mathematician could compute the expected payoff for any candidate, and determine the preference ordering on these expected payoffs.

⁹ Alternatively, one might presume that the agent assigns cardinal *utilities* to various states of affairs, and that her preferences regarding probabilistic chances of attaining states of affairs may be computed straightforwardly on the basis of these utilities. I do not use this formalism because it is not necessary for the points at issue here, and because I think it places unnecessary restrictions on what preferences an agent might have. E.g., I think an agent might rationally act on a brute preference for a *probable million with a slight risk of getting nothing over a probable mere thousand with a slight chance of getting an extra million*, even if this preference does not cohere in the standard ways with whatever cardinal utilities (if any) she assigns to receiving various dollar amounts *for sure*. However, I should stress that it is just as open to DBDT to use either formalism as it is to CDT.

Given the above definitions, both CDT and DBDT will say that a candidate C is *rationally permissible* for an agent with beliefs and desires as described in scenario description S just in case there is no other candidate D such that the expected payoff of D if adopted at the critical point in S^* would be preferable to the expected payoff of C if adopted at that critical point. For CDT, the candidates in question are choices, so this comparison of candidates will directly yield a conclusion about which choices are rationally permissible. For DBDT, the candidates in question are *dispositions*, so we need a small further step to reach a conclusion about rationally permissible *choices*: a *rationally permissible choice* is any choice that could be produced by a rationally permissible disposition.

In the case of DBDT, these definitions may be intuitively motivated by considering the two jobs (identified in section 2 above) that we use our rationality-talk to do. First, there is the job of preparing ourselves, people we care about, and especially our children to make successful decisions. If C and D are candidate dispositions to choose in response to the same scenario description, and C 's expected payoff in the described scenario is preferable to D 's, then we would do better to designate C , rather than D , as ‘rational’ in order to best do the work of preparing people for such scenarios. Similarly, saying C is rational would also better sustain the second job of rationality-talk, designating agents who are likely to do well in their scenarios, who are likely to give helpful advice regarding such scenarios, and who are likely to handle their friends’ investments well.

Let us now confirm that this presentation of CDT gives the treatment of Newcomb’s problem that CDT’s advocates intend. For CDT, there are two candidates to compare: choosing 1-box vs. choosing 2-boxes. We simplify things by moving to the related scenario description S^* which brackets

the details about the predictor, and says just that there is probability p that the opaque box has a million dollars in it rather than being empty (where p is determined by the agent's beliefs about what the predictor probably did). The expected payoff of choosing one-box in S^* involves a probability p of getting a million and otherwise nothing. The expected payoff of choosing two-boxes in S^* involves a probability p of getting a million plus a thousand, and otherwise just a thousand. Regardless of what value p has, an agent whose only relevant desire is to accumulate dollars must prefer the latter expected payoff to the former. Hence, CDT concludes, one should choose 2-boxes in Newcomb's problem.¹⁰

¹⁰ Notice that our moving to S^* served to make clear what role the probability p plays in determining (probabilistic) expected payoffs, where these expected payoffs were then compared in terms of their preferredness. The effective order of operations on my formalization of CDT is first to introduce the probability p , and then to worry about how preferences apply to *probabilistic* expected payoffs.

Standard formalizations of CDT reverse this order of operations. They begin by supposing that the agent's desires assign utilities to *non-probabilistic* outcomes (like getting a million dollars *for sure*). They then introduce the probability p into a picture that already includes these simple preferential elements. E.g., they would have us compute a utility for one-boxing by summing (a) the utility of one-boxing when the box is full, multiplied by p , together with (b) the utility of one-boxing when the box is empty, multiplied by $(1-p)$.

Given the common assumptions about utilities that these formalizations of CDT employ, it doesn't matter which order we do these operations in. (The utility of a probabilistic lottery over simple outcomes is equal to the sum of the utilities of the simple outcomes each discounted by their probability.) So long as these assumptions are in place, my formalization of CDT gives the same results as theirs. However, my formalization may be preferable because it generalizes naturally (and in a way that fits the intuitive motivation behind CDT) to cases in which an agent's desires do not meet the standard restrictive assumptions regarding utilities – cases about which standard formalizations of CDT have nothing to say.

Now let us apply DBDT to Newcomb's problem. There are two candidate dispositions to consider: one-boxism and two-boxism.¹¹ The critical point in Newcomb's problem comes prior to the visit to the predictor. The expected payoff of adopting one-boxism at this critical point is probably a million dollars (with a slight risk of unluckily getting nothing, given the predictor's fallibility), whereas the expected payoff of two-boxism is probably a thousand (with a slight chance of luckily getting a million plus a thousand). So long as the predictor is reasonably accurate, an agent who desires only to accumulate dollars must prefer the former expected payoff to the latter. Hence, according to DBDT, the rational disposition is one-boxism, and hence, the rational choice in Newcomb's problem is one-box.

We may summarize the key differences and similarities between DBDT and CDT as follows: CDT evaluates various *choices* by considering which outcomes those choices would yield *if made*; DBDT evaluates various *dispositions* by considering which outcomes those dispositions would yield *if adopted at the critical point*; and both approaches 'hold fixed' all factors outside the causal influence of the candidates being evaluated.

Because of the deep parallels between DBDT and CDT, the technical challenges that face DBDT closely correspond to those that face CDT. Both approaches must somehow account for ways in which an agent might represent and reason about the causal structure of her current scenario; both must somehow account for how she may 'hold fixed' factors outside

¹¹ I will ignore the complexities posed by mixed and probabilistic strategies. Either the predictor will see through these or else the predictor will not be confident of your one-boxing, and hence (by my description of the case) will reward you with an empty box. Either way, you would do better with a *resolute* strategy.

the causal influence of what is being evaluated; and both must somehow account for how the resulting expected outcomes may be compared on the basis of her preferences or ‘utilities’. Causal Decision Theorists have given a great deal of attention to these difficult issues.¹² I will not attempt to review or evaluate their proposals here, but I will note that whatever solutions CDT can find to these challenges will equally be solutions for DBDT (and, of course, vice versa).

Hence, we see that, from a formal perspective, CDT and DBDT are exactly on a par. It follows that purely formal methodology cannot serve to discriminate between these two theories or the notions of rational choice that they produce. Instead we must ask which of these two *formally pretty-well-defined* notions better deserves to be called a notion of ‘rationality’. We will take up this question in the final section. But first I will head off two lines of objection that might seem to press against DBDT.

6. THE SMOKING-CANCER PROBLEM

A common objection to the claim that one-boxing is rational involves an appeal to temporally extended cases that are structurally analogous to Newcomb’s problem. A representative example is the smoking-cancer problem.¹³

The Smoking-Cancer Problem.¹⁴ You must choose either to smoke or to abstain. You assign probability p to the claim that you have a certain gene – a gene which, in early childhood, causes its bearers to be likely both to be disposed to smoke in circumstances like your present ones and to get a horrible cancer later in life. You think it equally probable that you instead have an innocuous allele – an allele that makes its bearers unlikely to get cancer but also unlikely to be disposed to smoke in circumstances like your present ones. Finally, you would slightly prefer the *actual effects* of smoking to the effects of abstaining, while you would greatly prefer not getting cancer to getting it.

The smoking-cancer problem is structurally quite similar to Newcomb’s problem (as is illustrated in Figure 2). However, virtually everyone agrees upon what you should do in this problem: since there’s nothing to be done about the cancer, you *ought to* choose to smoke and thereby make the best of a potentially very bad situation.¹⁵ Given the clear structural analogy between these cases, many theorists have been inclined to dismiss one-boxer intuitions regarding Newcomb’s problem as anomalous, and to conclude on the basis of temporally extended cases that CDT must be correct.¹⁶

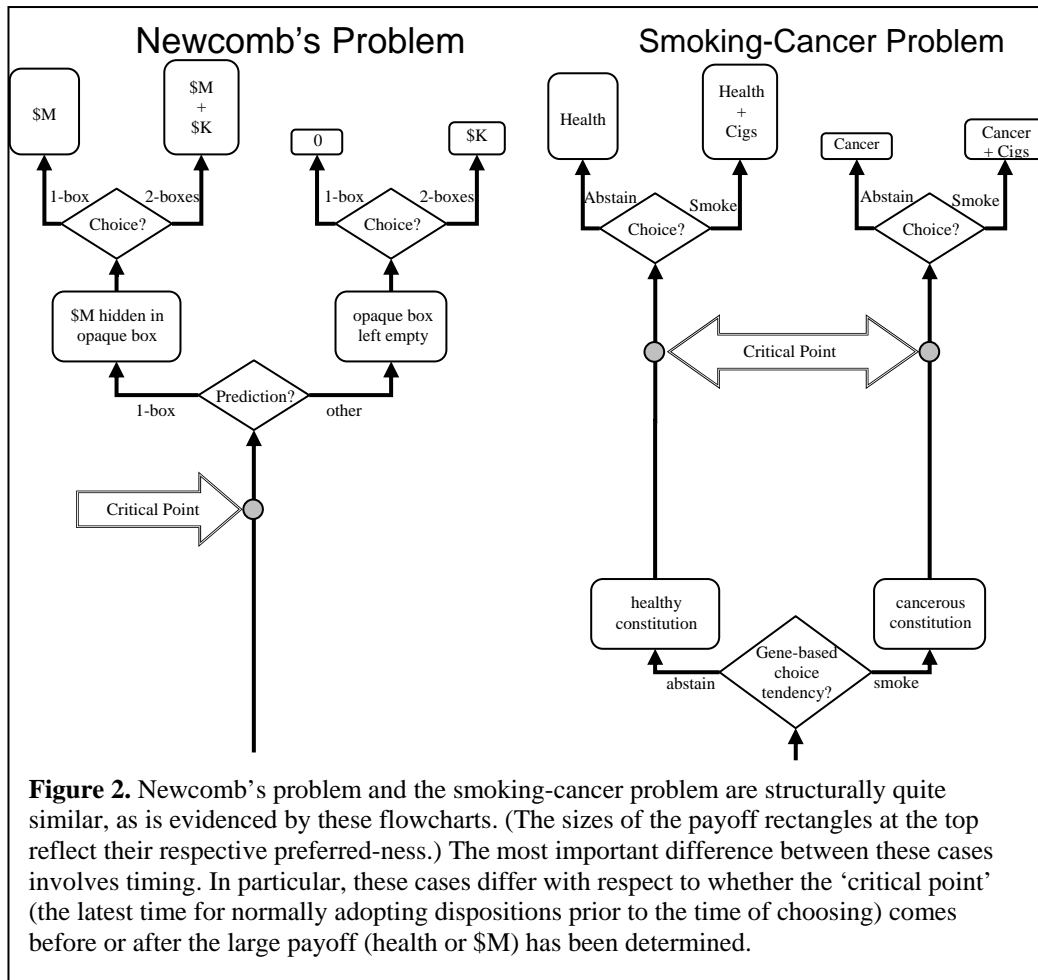
¹² E.g., Skyrms (1982), Lewis (1981a), Joyce (1999), Pollock (2002).

¹³ Nozick (1969) discusses a very similar case involving correlations between the disposition to choose academic vs. athletic occupations and the risk of getting a horrible disease. Gibbard & Harper (1985) discuss two cases in which childhood factors determine both a monarch’s choice-making dispositions and the length of his reign. Perhaps the most temporally extended Newcomb-like problem involves the Calvinist doctrine that, at the dawn of time, God preordained which select few will go to heaven based on divine foreknowledge of their moral decisions.

¹⁴ This problem is often associated with the statistician R.A.Fisher, who thought the best explanation for the observed correlation between smoking and cancer is that one gene causes both. A classic presentation of this problem, including its relation to Fisher’s work, is Levi (1985).

¹⁵ Whether one *can* choose to smoke is quite another matter. In discourse about rationality, *ought* clearly does not imply *can*. Rational norms say we all *ought to* reason in various ways, even if some of us – due perhaps to our genetic predispositions – are wholly incapable of doing this.

¹⁶ E.g., Pollock (2002, pg. 144) admits that “[t]he Newcomb problem itself commands conflicting intuitions” but, undaunted, goes on to stake his brief



My goal in this section is to show that DBDT both delivers the correct prescription in temporally extended cases like the smoking-cancer problem *and* makes clear why there is

positive argument for CDT solely on the basis of common intuitions about the smoking-cancer problem.

an evident intuitive difference between these cases and shorter cases like Newcomb's problem.

Recall that we defined the *critical point* of a given scenario description as the most recent time prior to the choice in question which would have been a natural opportunity for the normal shaping of dispositions. Disposition-shaping in humans is often a gradual social process. Hence, in decision scenarios that humans face, the critical point should typically be surrounded on both sides by ample time where the agent may engage in social interaction and learning.¹⁷ In Newcomb's problem, there is no such natural opportunity for normal disposition-shaping *after* the visit to the predictor, and prior to the choice, so the critical point must come before the contents of the opaque box are determined. In the smoking-cancer problem, there is ample time for normal disposition-shaping between the time when a child's risk of cancer is determined by the genes she is born with and the time when she makes the choice.¹⁸ Hence the critical point in this problem comes well after the risk of cancer is determined. (These critical points are also illustrated in Figure 2.)

¹⁷ It is a vague matter just how long before the choice is the appropriate time for dispositions to be evaluated, and just how much time is 'ample time'. For example, we may generate a *sorites* sequence by gradually extending the length of Newcomb's problem until even a staunch one-boxer's intuitions become murky. The borderline cases are intuitively problematic anyway (and, more to the point, they are not the sort of cases that we commonly use 'rationality'-talk to talk about) so it should not weigh against DBDT that it dictates no precise borderline.

¹⁸ For the case to work out as described, it must be that, in fact, normal disposition-shaping does not very often successfully run counter to the genetically predetermined tendencies to choose smoking or abstaining. This does not detract from the fact that there are normal times for the shaping of our dispositions.

This difference in critical points makes a difference in the prescriptions that DBDT gives. Since the critical point for Newcomb's problem is prior to the prediction, one-boxism is worth recommending as 'rational', for this recommendation would yield fruitful rewards if taken to heart at the critical point. Since the critical point in the smoking-cancer problem comes well after the agent's risk of cancer is fully determined, there is no reason to recommend abstinence as 'rational' at that late stage in the scenario.¹⁹ We should instead recommend smoking as 'rational', for these recommendations, if taken to heart, would yield the (admittedly small) pleasure of smoking.

Hence, DBDT says smoking is rational even while two-boxing is not, despite the structural similarity between the two cases. And hence, our commonly held, clear intuitions that smoking would be rational do not weigh against DBDT.

7. PARFIT'S ALLEGED COUNTER-EXAMPLES

Derek Parfit (1984, 2001) has attempted to give counter-examples to a disposition-based account of rational choice. One of his alleged counterexamples involves a predictor/extortionist who sometimes issues an irrevocable threat to kill an agent with a bomb unless the agent forks over a coconut. Parfit stipulates that the predictor/extortionist *probably* would not issue such threats to an agent that it recognizes is disposed to ignore them; but he also stipulates that there is a non-negligible chance that the

predictor/extortionist will deliver an irrevocable death-threat *regardless* of what dispositions an agent has.

Parfit suggests (correctly) that the clearly rational thing to do when faced with such an irrevocable death threat is to fork over a coconut. Parfit also suggests (incorrectly) that a disposition-based account is committed to saying something different. Given Parfit's stipulations and ordinary preferences, the expected payoff of a threat-ignoring disposition (a non-negligible risk of death) is intuitively *far worse* than the expected payoff of a capitulative disposition (the probable loss of a mere coconut). Hence, DBDT *agrees* with Parfit that, given the risk of death, it is rational to (be disposed to) fork over the coconut to the bomber. And hence Parfit's case fails to provide any intuitive weight against DBDT.

However, there are other predictor/extortionist cases upon which CDT and DBDT do disagree.²⁰ Imagine a crooked tax auditor who sometimes irrevocably threatens to initiate a mutually expensive audit unless given a small bribe; where it is

¹⁹ In the formalism described above, the historical details about how the cancer was determined are bracketed off, and re-construed as a fixed probabilistic constraints in S^* . The agent has probability p of having cancer and probability $(1-p)$ of not having it, and her adopting of one disposition rather than another cannot change that. Once these historical details are bracketed off, DBDT's treatment of the smoking-cancer problem exactly parallels CDT's treatment of Newcomb's problem.

²⁰ Parfit also discusses a case in which the predictor/extortionist is prepared to threaten certain death as a way to get the agent to open a locked safe. However, Parfit stipulates, even if the safe is opened, the extortionist will likely deal death anyway to lessen his risk of being identified later. In this scenario, a capitulative disposition would likely result in the agent's death; while an ignore-threats-and-avert-your-eyes disposition would be recognized, prompting the predictor/extortionist *probably* to flee empty-handed, and incurring only a slight risk of death. (Parfit tendentiously labels the latter sort of disposition 'the disposition to behave irrationally' – but this label begs the question against a disposition-based account.) I agree with Parfit that an agent might rationally adopt the latter sort of disposition before hand. Our only point of disagreement involves whether *acting on* this disposition would be rational – I say it would be; Parfit presumably says it wouldn't (so long as capitulating when threatened causes it to be less likely that death will be dealt). This point of difference mirrors our point of difference regarding the crooked auditor case I describe in the main text, but I think the crooked auditor case is more intuitively clear.

well-known that the delivery of such threats is usually – but not always – caused by the auditor’s having recently detected that the victim is disposed to capitulate.

In this case a threat-ignoring disposition may be worth aspiring towards because of the fact that (1) this disposition usually helps an agent avoid being exploited in the relevant scenario, and despite the fact that (2) this disposition compares poorly to other dispositions with respect to the consequences that are (occasionally) caused by its *becoming manifest as a choice* in such scenarios. CDT holds that only (2) is relevant to determining which choices (and hence which dispositions) are rational; while DBDT holds that both (1) and (2) are relevant to determining which dispositions (and hence which choices) are rational. Hence, CDT holds that it is rational to capitulate whenever the auditor threatens, even though capitulators predictably do less well on average than non-capitulators. Conversely, DBDT holds that it is rational always to ignore the auditor’s threats, even though threat-ignorers are disposed to pass up occasional (apparent) opportunities to reduce their losses. This intuitive split closely mirrors the intuitive split regarding Newcomb’s problem and hence offers no new reason to prefer CDT to DBDT.

8. WHICH DECISION THEORY SHOULD WE EMBRACE?

In Section 2, I advertised the general form of my argument for DBDT. I identified two important jobs that we use our rationality-talk for and suggested that we should embrace whichever decision theory produces a notion of rationality which is best capable of doing these jobs. In this final section, I drive home this argument.

The first important job we use rationality-talk for is to identify ways of making decisions that are worth adopting so

that we, and the people we care about, can become successful decision-makers. Naturally, DBDT’s notion of rationality is well-suited for this job, as this notion was explicitly geared towards designating as ‘rational’ those dispositions whose adoption would lead one to do well in the decision-making scenarios in which those dispositions may become manifest.

CDT’s notion of rationality does not do so well at this job. For example, CDT labels two-boxing as ‘rational’, even though all parties (including Causal Decision Theorists²¹) agree that we should seek to make ourselves and our children adopt one-boxing dispositions, as doing so would cause us to improve our future performance in Newcomb scenarios. When CDT tells you a choice is ‘rational’, you must then ask, “Yes, but is it the sort of choice that I should aspire to make, and that I should try to get my children to learn to make?” We use attributions of ‘rationality’ to *answer* questions like this, not to leave them wide open. This indicates that CDT’s notion of rationality cannot do all the work we expect a notion of rationality to do.

The second important job of our rationality-talk was to assess an agent’s decision-making capabilities, so that we might know whether to ask her for advice, or know whether to invest our money with her, or know how successful to expect her ventures to be. Once again, DBDT easily supports such usage. Not so CDT. When CDT tells you a trusted friend is rational, you must then ask, “Yes, but is she a good person to ask for decision-making advice? Is she a good person to invest my money with? Is she likely to achieve high payoffs in her ventures?” Attributions of ‘rationality’ are supposed to help answer these questions, not leave them open. So, again, this

²¹ For similar consideration of CDT’s advocacy of (what it says are) ‘irrational’ dispositions, see Parfit (1984) and Joyce (1999, pp. 153-4).

indicates that CDT's notion of rationality isn't really up for the job that we commonly call upon a notion of rationality to do.

A common response to considerations like these is to hold that certain scenarios 'punish rationality' and/or 'reward irrationality', and to hold that sometimes rationality isn't a desirable characteristic in a trusted advice-giver or investment handler. Of course, this response is correct, given CDT's quite peculiar notion of rationality. However, these are very *unfortunate* things for a 'rationality'-theorist to have to say – it is very unfortunate for a 'rationality'-theorist to have to concede that her notion of rationality can't really do the work we commonly call upon our concept of rationality to do. If some rival theory of rationality has to say such things less often,²² then this rival has a strong advantage over CDT. I claim such an advantage for DBDT. This constitutes a strong *prima facie* reason to accept the conclusion that, henceforth, we should embrace DBDT's notion of rationality.

At this point, there are three responses available to CDT. Response #1 would be to contest the *prima facie* argument just given, either by denying that we actually use the term 'rational' for the jobs that I've said we use it for, or by denying that DBDT's notion of rationality actually does better than CDT's at these jobs. This response doesn't seem at all promising. Response #2 would be to identify further jobs that we use the term 'rational' for, and to argue that CDT's notion of rationality is better suited to these other jobs than is DBDT's. If successful, this response would likely lead to the

²² Sometimes saying such things is unavoidable. Take, for example, the smoking-cancer problem. Sadly, most agents who end up smoking in this problem end up that way only under the double-edged aegis of the cancer gene. So, it turns out that rational behavior here is correlated with horrible outcomes. However, this 'punishment' is meted out *not* as a consequence of the smokers' mature dispositions to choose, and hence it does not reflect poorly upon the rationality of these dispositions.

conclusion that there are several useful notions in the ballpark of our concept of rationality and that it is a somewhat arbitrary choice which of these notions we will use the term 'rational' for. Response #3 would be to claim that there are independent reasons to embrace CDT's notion of rationality despite its failure to do the work that we commonly use the term 'rational' for. Let us briefly consider these latter two options in reverse order.

To give Response #3, the Causal Decision Theorist must produce some reason to embrace CDT's notion of rationality, independent of its capacity (or rather *incapacity*) to do the jobs that we commonly use a notion of rationality to do. I know of only one independent reason that is at all plausible – namely that many of us have strong pre-theoretic intuitions that favor CDT's prescriptions and even CDT's explanation for these prescriptions. For example, many people find the causal dominance argument (presented in section 3) for two-boxing in Newcomb's problem intuitively compelling. One might argue that the strength of these intuitions is itself a good reason to embrace CDT's notion of rationality, regardless of how poorly that notion does at the jobs that we commonly use our 'rationality'-talk for.

One obvious problem with this argument is that pre-theoretic intuitions do not point univocally in the direction of CDT. While many common intuitions favor CDT, there are also common and persistent intuitions that instead favor one-boxing, cooperating, and ignoring the crooked auditor's threats. Lacking agreement between these pre-theoretic intuitions,²³ it's

²³ Perhaps the strongest intuitive case for CDT involves the modified Newcomb scenario where both boxes are transparent. CDT (as well as its erstwhile rival EDT) advocates two-boxing in this scenario, for, at the point of the choice, the agent can tell precisely how much money each option will yield. However, one-box dispositions get higher average payoffs in this

not at all clear how these intuitions might push us to accept one theory rather than the other.

In the face of this intuitive standoff, many Causal Decision Theorists are inclined to call into question the genesis and the credentials of their opponents' intuitions. I concede to Causal Decision Theorists that many of their opponents' intuitions probably are sustained, in large part, just by the common success of a simple *evidentialist heuristic*:

Evidentialist Heuristic: Usually, one rationally ought to make a given choice if that choice is *correlated with* a payoff better than that of any alternative choice.

Causal Decision Theorists are happy to accuse their opponents of naively over-generalizing this heuristic to cases (like Newcomb's problem) in which we have no obvious reason to think it should apply.²⁴ I agree that once we've thus explained the genesis of one-boxer intuitions in a way that doesn't presuppose their *general* truth, the presence and persistence of these intuitions is not *itself* a reason to favor one-boxing theories like DBDT.

But what most Causal Decision Theorists have not noticed is that pre-theoretic *two-boxer* intuitions have the same questionable status as pre-theoretic one-boxer intuitions. For

scenario, and they also are most worth aspiring towards – any loving parent should teach her child to be the kind of person who would one-box in this scenario. Thus, DBDT holds that one-boxing is rational even in this case. I acknowledge that this conclusion runs counter to some commonly held intuitions; but still I maintain that this conclusion is correct, and that these intuitions are misplaced. Since we are committed to using the term 'rational' to designate the ways of choosing that we think our children should adopt and that we think would tend to make an investment handler do well, we are therefore committed to saying that one-boxing is 'rational' even in this version of the Newcomb problem.

²⁴ See, for example, Gibbard & Harper (1985, pp. 152-3).

two-boxer intuitions may themselves be explained as a naïve generalization of the causal dominance heuristic presented in section 3:

Causal Dominance Heuristic: Usually, one rationally ought to make a given choice if, regardless of how things outside one's control turn out to be, one would be better off if one made that choice than one would be if one made the other choice.²⁵

Like the *evidentialist* heuristic, this causal dominance heuristic works perfectly well in most everyday cases. But, as with the evidentialist heuristic, we have no obvious, pre-theoretic reason to think the causal dominance heuristic should also apply in less ordinary cases like Newcomb's problem. We may explain the genesis and persistence of two-boxer intuitions just by noting the common usefulness of the causal dominance heuristic. Once we've done this, the evidential status of two-boxer intuitions is no different from the evidential status of one-boxer intuitions. We have no good pre-theoretic reason to think that any of these intuitions should be a good guide to a *generally* interesting and useful theory of rationality.

The ultimate arbiter in this debate *cannot* be pre-theoretic intuitions, for such intuitions do nothing more than reflect the heuristics that work well in everyday cases, and we

²⁵ By 'heuristic' I mean an easily applied rule that we commonly employ in our reasoning, even though its not absolutely clear that this rule always gives the right answers – i.e., even though this rule is intuitively defeasible. The intuitive defeasibility of the evidentialist and causal dominance heuristics is evidenced by the fact each of these is suspected to be wrong by some common intuitions about Newcomb's problem. In calling a rule a heuristic, I beg no questions regarding whether or not that rule might turn out to play a central role in the theory we end up adopting. E.g., I think 'promote the greatest overall happiness' is *clearly* a heuristic, but, for all I know, it might also be a central tenet of the correct moral theory.

have no good reason to think that these (often conflicting) heuristics should apply in less ordinary cases. Even if it were possible to come up with a general theory of rationality that optimally conforms to all our pre-theoretic intuitions, there still would be no guarantee that such a theory would be generally interesting or useful. Instead, it seems, our choice of a theory of rationality must be guided by judging how well the various candidates can do at the jobs that we are going to use a theory of rationality for. Hence, I conclude, Response #3 will not work.

The only remaining option for CDT is Response #2 – to identify some further jobs that we want to use rationality-talk for, and to argue that CDT’s notion of rationality can do better at these jobs than does my disposition-based notion. I can’t offer a principled argument saying that there is *no* practical advantage for CDT’s notion of rationality – that there is *no* interesting work that we could use CDT’s notion of rationality for, but couldn’t use DBDT’s notion for instead. However, I also can’t think of any such advantage myself. Instead, I will leave this as a challenge for the supporters of CDT. One challenge to CDT has already been posed many times: “If you’re so rational, why aincha rich?”²⁶ To this I add, “If you’re so rational, what good does knowing that do any of us?” or, more glibly, “If you’re so rational, who cares?”

Unless the Causal Decision Theorist can meet this challenge, we will be forced to conclude that CDT is nothing more than a sometimes-useful heuristic and a formal framework of merely academic interest. CDT is the result of pushing an attractive intuition far beyond its realm of usefulness, and this result is clearly not capable of doing the work that we commonly expect a theory of rationality to do. It is time to lay CDT aside, and to return to the hard work of

developing a generally useful theory of rational decision-making. As a first step, we should move away from the traditional choice-based conception of rationality and explore the disposition-based decision theory laid out above.

ACKNOWLEDGMENTS

This work was supported by NSF grant no. IRI-IIS-0080888. I am indebted to helpful comments from Travis Fisher, Allen Habib, Alan Hájek, Terry Horgan, James Joyce, John Pollock, Kenneth Presting, David Schmitz, Joseph Tolliver, Sarah Wright, and audiences at the University of Arizona, the Society for Exact Philosophy, and the Australian National University.

REFERENCES

- Campbell, R. and L. Sowden, eds (1985): *Paradoxes of Rationality and Cooperation: Prisoner’s Dilemma and Newcomb’s Problem*. Vancouver: UBC press.
- Gauthier, D. (1985): ‘Maximization Constrained: The Rationality of Cooperation’, in Campbell & Sowden (1985, pp. 75-93).
- Gauthier, D. (1986): *Morals by Agreement*. Oxford: Clarendon.
- Gibbard, A. and W. Harper. (1985): ‘Counterfactuals and Two Kinds of Expected Utility’, in Campbell & Sowden (1985, pp. 133-58).
- Jeffrey, R. (1983): *The Logic of Decision*, 2nd edition. Chicago: University of Chicago Press.
- Joyce, J. (1999): *The Foundations of Causal Decision Theory*. Cambridge UP.
- Levi, I. (1985): ‘Common Causes, Smoking, and Lung Cancer’, in Campbell & Sowden (1985, pp. 234-47).

²⁶ For example, this challenge appears in the title of Lewis (1981b).

- Lewis, D. (1979): 'Counterfactual Dependence and Time's Arrow', *Noûs*, 13, 455-76.
- Lewis, D. (1981a): 'Causal Decision Theory', *Australasian Journal of Philosophy* 59, 5-30.
- Lewis, D. (1981b): 'Why Ain'cha Rich?', *Nous* 15, 377-80.
- Nozick, R. (1969): 'Newcomb's Problem and Two Principles of Choice', *Canadian Journal of Philosophy* 7. pp. 213-225. Reprinted in Campbell & Sowden (1985, pp. 107-132).
- Parfit, D. (1984): *Reasons and Persons*, Oxford: Oxford UP.
- Parfit, D. 2001. 'Bombs or Coconuts or Rational Irrationality', in *Practical Rationality and Preference: Essays for David Gauthier*. Christopher W. Morris and Arthur Ripstein, eds. Cambridge UP. pp. 81-97.
- Pollock, J. (2002): 'Causal probability', *Synthese* 132, 143-85.
- Skyrms, B. (1982): 'Causal Decision Theory', *Journal of Philosophy* 79, 695-711.

Department of Philosophy
Southern Methodist University
E-mail: fisher@smu.edu