Newcomb's Problem as Cognitive Illusion

Peter Slezak (p.slezak@unsw.edu.au)
Program in Cognitive Science
School of History & Philosophy of Science
University of New South Wales
Sydney, NSW 2052, Australia

Abstract

Richard Jeffrey (1983) has said that Newcomb's Problem may be seen as a rock on which Bayesianism must founder. Despite a vast literature of great technical subtlety and complexity, no solution has emerged. Most recently, Jeffrey (2004) renounced his earlier position and no longer regards Newcomb's problem as a genuine decision problem at all. Rather, Jeffrey suggests, "Newcomb problems are like Escher's famous staircase on which an unbroken ascent takes you back where you started." Jeffrey's analogy is apt for a puzzle whose specific logical features can be precisely articulated. I offer a novel analysis of the problem going beyond mere analogy to reveal the source of its persistent intractability. In the spirit of Jeffrey's analogy, I propose that the central problem has arisen from the misguided attempt to reconcile the impossible science-fictional features of the story with a plausible causal analysis and, therefore, the contest between causal and evidential decision theories is misconceived. As the analogy with Escher figures suggests, Newcomb's Problem is an instance of a compelling cognitive illusion whose underlying mechanism is proposed here.

Keywords: Newcomb's Problem; decision theory; rational choice; Bayesianism; paradoxes of self-reference.

Introduction

The Problem involves a choice between two alternatives: Of two boxes A and B, you may choose either to take Box B only, or you may choose to take both boxes A and B. Box A is transparent and contains $1,000; Box B is opaque and contains either a million dollars or nothing, depending on the prediction of the demon who places the money there. If the demon predicts you will choose only Box B, then he will place the million dollars in it. If he predicts that you will choose both boxes, he will leave Box B empty. This predictor demon is known from previous experience to be extremely reliable, making correct predictions 95 percent of the time. He makes his prediction, and depending on what he predicts about your choice, either places the million dollars in Box B or not. He departs and can no longer influence the outcome, and then you make your choice. What do you do?

Given the high reliability of the demon's predictions, the principle of subjective expected utility recommends taking only Box B since there is almost certainty of winning a million dollars. However, since the demon either places the money or not prior to your choice and can no longer influence the situation, the principle of dominance recommends taking both boxes since you will be $1,000 better off regardless of what the demon has done. There is no point leaving a certain gain of $1,000 when it can not influence the outcome of the choice.

Intellectual Cripples?

Newcomb's Problem may suggest ill-understood features of rational choice behavior and appears to reveal anomalies in our tacit principles of decision making like the tradition of research on 'heuristics and biases' (Tversky and Kahneman 1974). On such a view, Newcomb's problem is an addition to the list of such famous "paradoxes" as those of Allais (1953) and Ellsberg (1961). On standard accounts, Newcomb's Problem appears to support Slovic's jaundiced view that research into such phenomena "has led to the sobering conclusion that, in the face of uncertainty, man may be an intellectual cripple, whose intuitive judgements and decisions violate many of the fundamental principles of optimal behaviour" (Wright 1984: 114). I will argue that such pessimism is unwarranted, in this case at least, because the problem is merely a pseudo-problem and, therefore, needing not to be solved but to be dissolved.

Two decades after having introduced the puzzle, Nozick's (1993) judgement that no resolution has been completely convincing suggests that something essential has been overlooked. As the few 'no-box' accounts suggest, a radically new approach is needed that can meet the important criterion of adequacy on any solution – namely, revealing the source of the problem's peculiar obduracy. In particular, I will suggest that, contrary to the nearly universal view, Newcomb's Problem does not raise any questions concerning rationality or decision theory. However, if this solution in the spirit of Jeffrey's apostasy is good news about our capacity for rational choice, it is bad news about other cognitive abilities of interest. Revealed as an instance of a familiar class of paradoxes, Newcomb's Problem is a manifestation of problems that have plagued theorizing about the mind. In the extensive literature on Newcomb's Problem, Sorensen (1987, 1988), Priest (2002) and Maitzen and Wilson (2003) are among the few who appear to have noted the affinity with a family of problems having nothing essentially to do with rationality or decision theory. I show that Jeffrey's late insight can be given precise formulation which reveals unnoticed, but not entirely coincidental, similarity between Newcomb's demon and that of Descartes: Just as Descartes' demon systematically thwarts our beliefs, so Newcomb's demon systematically thwarts our choices.

Goofball Cases

Zeno's paradox of Achilles and the Tortoise is not resolved by trying to reconcile its conclusions with algebraic calculation of their relative positions over time. The conclusion that Achilles cannot overtake the tortoise is taken as a *reductio ad absurdum* of Zeno's argument, and the intellectual task is to expose its fatal flaw. Newcomb's problem has not been approached in this way, since the intellectual effort has been expended mainly on reconciling the famous scenario of the predicting demon with some plausible causal structure (Eells 1982). Thirty years of failure has not been widely taken as evidence of the futility of this project and the underlying conception of the project. Ironically, however, taking the puzzle seriously entails embracing the paradox rather than attempting to re-tell the story in a way that avoids it.

David Lewis (1979) notes that some have dismissed Newcomb's Problem as a "goofball" case unworthy of serious attention. On the contrary, however, I suggest that it is worth of serious attention *because* it is a goofball case of a specific kind. In the literature, illuminating similarities have been noted between Newcomb's Problem and various structurally similar problems such as "common cause" cases and Prisoner's Dilemma. However, these real-life analogs of Newcomb's Problem have been crucially misleading by diverting attention from the dis-analogy and the essential function of the science-fiction of a demon predictor. Neglecting the implications of the supposed supernatural powers has led to missing their precise role in generating the recalcitrant perplexity.

The Shadowy Predictor

A central feature of Newcomb's Problem is the peculiarity of the apparent link between one's choice and the previously determined contents of the opaque Box B. Causal theorists such as Gibbard and Harper (1978) propose simply ignoring the link and recommend the 'two-box' solution as rational despite being forced to admit that you will fare worse in choosing it. They explain "We take the moral of the paradox to be … If someone is very good at predicting behavior and rewards predicted irrationality richly, the irrationality will be richly rewarded" (1978: 369). However, if "irrationality" so-called is richly rewarded, it must be rational to act in such ways. This "solution" has a distinct air of question-begging and appears "to just repeat one of the arguments, loudly and slowly" as Nozick (1969) originally remarked. Gibbard and Harper are not alone. Recently, McKay (2004) emphasizes the predictor's unfathomable ability and suggests that his reliability is so extraordinary "that it undermines your belief that your choice can have no causal influence" and even "challenges the conviction that the action of the predictor is genuinely in the past" (2004: 188). McKay says that faced with the predictor's reliability, "it is not impossible that you would come to believe that there is some very cleverly arranged cheating going on" (2004: 188). Indeed, *if it were not science fiction* but a real case, we would be desperate to find some plausible basis for the phenomenon. Equally, if we encountered an Escher staircase in real life, we would be anxious to resolve the anomaly in a way that is consistent with geometry and physics, as Richard Gregory (1981: 409) has demonstrated with an apparent real-life impossible Penrose Triangle. However, McKay's analysis leads us astray, like most others, because she fails to take seriously the fiction of an inherently occult "acausal synchronicity" between our choices and the predictor's actions. McKay is not at liberty to re-tell the story in a way that eliminates the puzzle arising from the predictor's supernatural ability.

The peculiarity of the apparent link between one's choice and the previously determined contents of the opaque box is the central, defining feature of Newcomb's Problem. It is this mysterious a-causal link that prompted Jeffrey's (1983: 25) original characterization of the problem as "a secular, sci-fi successor to the problems of predestination." Thus, the science-fictional nature of the problem frees, indeed *precludes*, us from the need to wonder about *how* such a predictor could possibly accomplish his success.

Common Causes

Burgess (2004) also defends the causalist position, bringing into relief the same difficulties. Burgess argues that Newcomb's Problem must be understood as a 'common cause' problem following Eells (1982). On this basis he argues "the evidence unequivocally supports two-boxing as the rational option" (2004: 261). Burgess does not consider the possibility that there might be no right choice at all in principle – the 'no-box' alternative that has been independently raised by a few authors for varied reasons (Levi 1975, Sorensen 1987, Priest 2002, Maitzen and Wilson 2003). Again, we see a characteristic difficulty when Burgess is forced to count the lesser expectation of two-boxing as, nonetheless, "most desirable." At the very least, it is to strain ordinary usage to claim that greater monetary expectation is not necessarily the most desirable. Above all, assimilating Newcomb's Problem to common cause cases is to stack the deck and to prejudice the case in favor of two-boxing unless the differences are inessential. However, Newcomb's Problem can only be categorized as a case of common cause on certain untenable, question-begging assumptions.

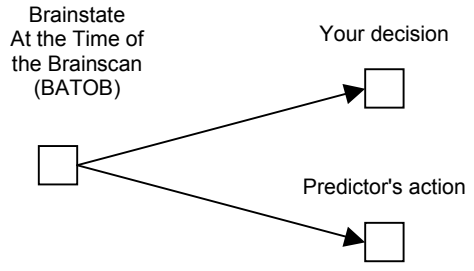

Figure 1.

Following Eells (1982), Burgess's picture may be represented in the schema of Figure 1, showing the analogy with the case in which there is a high correlation between smoking and cancer but smoking is assumed not to cause cancer because they are both assumed to be caused by a common gene. Despite providing unwelcome evidence that one has the cancer gene, smoking is the rational choice if it provides

pleasure – the analog of choosing two boxes in Newcomb's Problem. Since backwards causation is ruled out, these choices cannot affect the earlier facts for which they merely provide evidence. In Eells' terminology, 'symptomatic acts' and 'symptomatic outcomes' in such cases are highly correlated but causally independent. Eells (1982: 210) says that despite first appearances, the differences between such cases and Newcomb's Problem are not really important ones. He says "it seems that on any plausible account of successful prediction, the causal structure must be of this form" (1982: 210-11). However, the point is precisely that we are not necessarily required to seek any *plausible* account in this sense. As we will see, the effort to do so has not merely involved gratuitous inventions that go beyond the story's specifications, but has diverted attention from the specific source of puzzlement created by the science-fiction. Significantly, Eells says that the common cause structure is the only kind of causal explanation of the predictor's success that he can think of, since "it seems presupposed by much of our inductive reasoning that a high statistical correlation has a causal explanation" (1982: 210). The question that remains unasked is why we should search for any thinkable, plausible explanation of this kind consistent with our inductive reasoning rather than assume that the problem is inherently incapable of being reconciled with any causal structure. Closely analogous are those science-fiction stories of time travel whose impossibility according to physical theory creates inevitable, diverting paradoxes. If we are not misled by the analogy with problems that *are*, in fact, realizable, such as common cause cases, we may simply accept the occult relation between our choice and the predictor's earlier action, and thereby focus on the *logical* structure of the puzzle, rather than its supposed causal structure.

We see from Figure 1 that, on Burgess's account, the diagnostic Brainstate At the Time of Brainscan (BATOB) is assumed to be the common cause of your decision and also the Predictor's action. However, the *relevant* brainstate and the decision cannot be separated in this way. The Predictor's action cannot be caused by the relevant brainstate unless either (a) there is backward causation, contrary to the specifications of the problem, or (b) the brainstate in question is taken to be one preceding both the actual decision and also the Predictor's action. It is only by separating some diagnostic brainstate from the one directly responsible for (i.e. identical with) the decision that Newcomb's Problem can be characterized in terms of Eells' schema for common cause problems. That is, Burgess assumes that the Predictor bases his action on some brainstate earlier than the one actually constituting the decision itself, but this opens a questionable gap in the causal sequence between the brainstate and the decision. Of course, like the so-called precursor presentiment or 'tickle' (Eells 1984), this gap provides the room for supposing that the actual decision might somehow deviate from the one determined by the brainstate being scanned and thereby deviate from the demon's prediction. Burgess's strategy attempts to avoid the demon's prediction by assuming that a presentiment or tickle is merely evidence of the choice and not identical with the choice decision itself which is, thereby, assumed to be possible contrary to the tickle. Clearly, however, this must be ruled out since, *ex hypothesi*, as a reliable predictor, the demon will anticipate such a sneaky strategy. Although futile, this ruse captures something of the inescapable paradox of trying to avoid one's self – to flee one's fate. As we will see, the assumption of the demon introduces an inessential step in what is, in fact, the anticipation of one's own decision. If not through backward causation, the 'state of nature' is, nonetheless, not independent of my choice. Contrary to Burgess's strategy, if the brainstate relevant to the decision is taken to be the very one most directly responsible for (i.e. identical with) it, then it cannot be a cause of the Predictor's action unless it contravenes the stipulation prohibiting backward causation. Strictly speaking, the only brainstate that is relevant is the one at the moment of decision, precisely when it is impossible for you to incorporate it into any deliberation – for logical reasons, as we will see.

Appointment in Samarra

In a revealing remark, Burgess explains that Newcomb's problem "is taken to be a distinctive kind of common cause problem in that you are presently in a position to influence the nature of the common cause. Indeed, all you have to do to influence it appropriately is to make a commitment to one-boxing" (2004: 283). Burgess evidently hopes to distinguish the commitment to one-boxing from the actual, real choice itself, thereby contriving a means to avoid the predictor's supernatural powers. However, this attempt to split the commitment from the choice is clearly a futile attempt to outwit the Predictor by separating the diagnostic brainstate from the actual decision. Such an analysis is clearly ruled out by the specifications of the problem since the Predictor cannot be assumed to base his prediction on the wrong, or irrelevant, earlier diagnostic brainstate – Burgess's BATOB in Figure 1. Of course, without this spurious assumption, the parallel with common cause cases cannot be maintained. The attempt to outsmart the Predictor with a two-stage strategy is futile because it misconceives or reformulates the problem, thereby avoiding it rather than solving it. Burgess' suggestion that we switch commitments in order to trick the Predictor recalls the famous story told by W. Somerset Maugham in which a servant is frightened when encountering Death in the market place of Baghdad and, taking the master's horse, flees to Samarra. Recounting the meeting to the master, Death says: "I was astonished to see him in Baghdad, for I had an appointment with him tonight in Samarra."

Taking Science Fiction Seriously

Burgess accepts that the contest between causal and evidential decision theories is meaningful and defends the former by assimilating Newcomb's Problem to realizable 'common cause' cases where the recommendation of causal theory is generally acknowledged to be the rational one. Schmidt (1998), too, is concerned to "prevent this beautiful paradox from being classified as physical nonsense" (1998, 68) and seeks to refute claims that Newcomb's Problem is "physically nonsensical and impossible to imagine" (1998, 67). However, despite claiming that backward causation is realizable in a physically plausible way, Schmidt is concerned only with an agent's subjective impressions and intuitions

that constitute the problem and not its solution. It is striking to see that the effort to make sense of the puzzle within the framework of plausible scientific reasoning has persisted as the most widespread approach since Eells' (1982) account along these lines.

Schrödinger's Cash

The anomalous correlation of our choice with the contents of the opaque Box B has been illuminated by a creative resort to other suggestive mysteries. Although it should not require argument to explain why Wolf's (1981) appeal to Heisenberg uncertainty, superposition of quantum states and observer effects cannot be taken seriously in this context, it is a revealing manifestation of the same motivations and strategies we have already seen. Indeed, Wolf's analysis is not merely absurd or desperate, and its appeal to quantum effects is analogous to McKay's (2004) resort to extra information about unknown causal structure. Both seek to deal with the mysterious reliability of the Predictor by gratuitous, invented appeals beyond the actual specifications of the puzzle. In this case, the resort to quantum physics is no more unwarranted than the usual appeal to causal structures within classical physics. However, despite the manifest absurdity of invoking quantum effects, by taking the occult link between choice and box contents seriously, Wolf's analysis acknowledges and embraces the essential science-fictional feature of Newcomb's Problem that has been the source of its notorious recalcitrance. While Burgess sees the link as a realistic opportunity to "influence the nature of the common cause" (2004: 283), Wolf proposes in the same spirit that "the million dollars is in paradox-land where it is in the box and not in the box at the same time" (1981: 150). Like Schrödinger's Cat that is both dead and alive, the money is in a superposition of states, being both in the box and not in the box until you make your choice. Wolf says "Choosing both boxes creates [opaque] box B empty. Choosing box B creates it one million dollars fuller" (1981: 150). Although the failure to be convinced by Wolf's account need not be evidence of "our Western preconditioned minds" as he claims, Wolf's eccentric resort to an observer-induced effect captures the crucial peculiarity of the problem of uncaused backward correlation.

Twin Prisoner's Dilemma

Like Levi (1975, 1982), McKay blames the perplexity of Newcomb's Problem on the obscurity or under-specification of the choice conditions. On the contrary, however, I suggest that the problem is neither obscure nor ill-defined but rather clear, though formally paradoxical in a strict and familiar logical sense. This is a 'no-box' view according to which the very idea that there is a right choice is misconceived since the problem is ill-formed or incoherent in an identifiable way. Burgess rejects analyses of Newcomb's Problem as a variant of Prisoner's Dilemma (Nozick 1969, Lewis 1979), but acknowledges that Prisoner's Dilemma would be a case of common cause if the players were to make and keep a prior agreement to cooperate. However, Burgess misses the significance of the case of duplicates or 'twin' prisoners who are assumed to be identical. The strategic or normal form representation (Figure 2) shows your payoffs (your choices in rows, twin's choices in columns):

	2 Boxes	1 Box
2 Boxes	$1,000 + 0	$1,000 + $ 1M
1 Box	0	$ 1M

Figure 2.

Commentators appear not to have drawn an obvious consequence of this analogy. The isomorphism of the two problems reveals that Newcomb's Problem is a way of contriving a Prisoner's Dilemma against one's self. The other player in Newcomb's version of Prisoner's Dilemma is one's self mediated by the reversing mirror of the predicting demon. The illusion of a partner, albeit supernatural, disguises the game against one's self – the hidden self-referentiality that underlies Priest's (2002) diagnosis of "rational dilemma," Sorensen's (1987) instability and Maitzen and Wilson's (2003) vicious regress. Newcomb's Problem is a device for externalizing and reflecting one's own decisions.

The hidden circularity facing the decision-maker in Newcomb's Problem may be seen more directly by considering its formulation in extensive form (Figure 3). Here we may assume without material difference to the problem that the demon makes his decision following the agent's choice, though without knowing what the agent's move was.

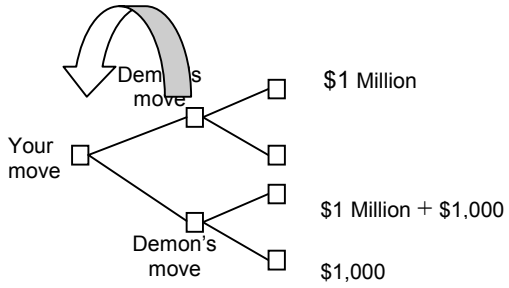

Figure 3.

As we contemplate our best move, we must consider the next level in the game tree representing the demon's decision which is actually reflection on the very same game tree and, in particular, the first node that we currently occupy. That is, the branches from the second level nodes are copies of the first node, since they represent the demon's prediction of our own decision at the first node. As we deliberate, we represent the demon's deliberations that incorporate our own.

The analogy between Newcomb's Problem and Prisoner's Dilemma appears to have obscured the indirect self-referentiality involved in the latter by highlighting the two-person structure of the contest. The other player in the Newcomb case is actually one's self mediated by the reversing mirror of the predicting demon. The predictor is not merely a fiction providing insufficient "extra information" about causal structure that might be reconciled with a meaningful choice as both McKay and Levi suggest. Rather,

the choice is paradoxical in a strict and familiar logical sense, thereby permitting a precise specification of the source of its notorious perplexity. It is along these lines that Sorensen's (1987) "instability," Priest's (2002) "rational dilemma" and Maitzen and Wilson's (2003) "hidden regress" have important affinities with the present analysis and promise to break the long-standing stalemate.

"Deliberation Crowds Out Prediction"

Levi (1997: 80) and Schick (1979) have drawn attention to the problem arising when a deliberating agent adopts the posture of spectator concerning his own performances. In such cases, Levi says the agent cannot coherently assign unconditional probabilities to hypotheses as to what he will do for "deliberation crowds out prediction," or as Schick puts it, "logic alone rules out our knowing the whole truth about ourselves" (1979: 243). However, neither Levi nor Schick appear to offer this analysis of self-referential paradox specifically as a diagnosis of Newcomb's Problem and the source of its perplexity.

Recently, Maitzen and Wilson (2003) suggest that the puzzle arises from an infinitely long, infinitely complex proposition and a hidden regress that is incomprehensible. They have an important insight, but their account is misleading because the regress they note is a symptom of a familiar circularity and paradox which is, nonetheless, perfectly intelligible. Significantly, Maitzen and Wilson offer the illustration of the Liar paradox, suggesting that "something can look comprehensible without being so." However, the analogy with the Liar is closer than Maitzen and Wilson themselves appear to think, and the moral to be drawn from it is quite different. Maitzen and Wilson suggest that "the classical Liar sentence makes trouble only because people mistakenly take it to mean something." They explain "Every constituent of the sentence is comprehensible, but, arguably, the sentence itself is not" (2003, 159). On the contrary, however, the classical problem of the Liar arises precisely because the sentence is perfectly meaningful, but appears to be both true and false. Indeed, the paradox with its contradictory truth values would not arise if the Liar sentence were meaningless. Thus, Maitzen and Wilson miss the way in which the Liar paradox does indeed illuminate Newcomb's Problem when its relevance is properly understood.

The logical source of the problem may be seen from a schematization that brings out its logic in a familiar form:

(a*) I choose ~ (b*)
[I choose the opposite of whatever the demon predicts]

(b*) The demon predicts (a*)
[The demon predicts whatever I choose]

Substituting appropriately, we get:

(a*) I choose ~ (The demon predicts (a*))

Assuming that the demon predicts reliably,

(a*) I choose ~ (a*)
[I choose the opposite of whatever I choose]

We see the analogy with the Liar Paradox and the family of related conundrums arising from self-reference. The Liar Paradox may be represented as:

(p*) ~ (p*)
[It is not the case that (p*)]

More generally, the problem arising from self-reference is a version of 'paradoxes of grounding' (Herzberger 1970). Slezak (1983) has shown that Descartes' notorious Cogito argument, too, may be captured in the following formula that says of itself that it is doubtful and is, therefore, certain:

(x*) I doubt (x*)

The sentence (x*) captures Descartes' insight, for attempting to doubt (x*) means considering it to be false. In turn, since (x*) asserts 'I doubt (x*)', its falsity means that I don't doubt (x*) or, in other words, that (x*) is indubitable or certain. This diagonal sentence makes perfect sense of remarks which otherwise remain obscure or irrelevant. Thus, Descartes (1984: 418) says "my doubt and my certainty did not relate to the same objects: my doubt applied only to things which existed outside me, whereas my certainty related to myself and my doubting."

The analogy of the decision proposition (a*) with the Liar sentence (p*) and the Cogito sentence (x*) is evident. As Schick noted in relation to decisions, "logic alone rules out our knowing the whole truth about ourselves," suggesting that Descartes' demon and Newcomb's are one and the same. The seemingly unrelated problems appear to reflect inherent cognitive mechanisms akin to familiar cognitive illusions – in this case arising in the attempt to understand one's self.

Experimental Realization

Schmidt (1998) claims that Newcomb's Problem may be realized in classical physics with backward causation, but he defends only a subjective ersatz – the impression or intuition of backward causation in an "anthropically oriented causal description" which is not the real thing that others have ruled out. Moreover, Schmidt's "strange but possible story" involving tiny sub-particle "dwarf" creatures whose scientific knowledge is millennia ahead of our own may be questioned on the grounds of plausibility. However, by exploiting available techniques and well-known facts of neuroscience, we may rig an experimental arrangement that simulates the predicting demon and presents exactly the same choice problem to a subject. Well-known work (Libet 1985) on the subjective delay of consciousness of intention, the so-called 'readiness potential' or 'preparatory response,' provides a means for a laboratory simulation of Newcomb's Problem. Instead of Schmidt's fiction of sub-particle dwarfs, we may obtain precise, reliable prediction of a subject's actions from the prior state of their brain, just as in the usual speculations (Burgess's BATOB), but without depending on utopian neuroscience. EEG recordings from scalp electrodes show that the instant of a subject's conscious decisions are around 400 milliseconds later than the onset of cerebral electrical

activity that is the substrate of voluntary action. It is a trivial matter to connect scalp electrodes to a computer screen in such a way that the readiness potential for a choice would cause a million dollars or nothing to be placed in the second opaque box before the subject consciously "makes the decision." This would amount to a prediction by the apparatus of the subject's choice. Exploiting readiness potentials is simply a way of making just those predictions about our actions from earlier brain-states that are standardly assumed in the literature. Of course, the interest of this case derives from the direct way that the experiment reveals what is implicitly involved in making choices under the circumstances of the problem – namely, the futile attempt to incorporate prediction of our own behaviour into our deliberations.

Conclusion

The seemingly unrelated problems we have noted appear to reflect certain deep, inherent cognitive mechanisms arising from attempting to understand one's self. Self-reference gives rise to well-known paradoxes in logic which may be abstract schemata capturing deep psychological processes. These cognitive mechanisms may be inherent features of our mental representations of the world insofar as they attempt to encompass themselves. Long forgotten in the philosophical literature, an analysis of self-knowledge along these lines was given by Josiah Royce (1900) in his Gifford Lectures, and more recently by K. Gunderson (1970) as an account of the aetiology of certain puzzles about the mind. Such reasoning may, after all, be seen as akin to the cognitive illusions uncovered in the 'biases and heuristics' research program, though in this case of a particular, limited variety. Happily the domain of thought affected seems to be narrowly confined: The cognitive illusions in question appear to violate the norms of rational thought only in philosophical speculation.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque. *Econometrica*, 21, 503-546.

Burgess, S. (2004). The Newcomb Problem: An Unqualified Resolution, *Synthese* 138, 261-287.

Descartes, R. (1984). The Search for Truth, *Philosophical Writings of Descartes II,* J. Cottingham, R. Stoothoff & D. Murdoch. Cambridge University Press.

Eells, E. (1982). *Rational Decision and Causality*. Cambridge: Cambridge University Press.

Eells, E. (1984). Metatickles and the Dynamics of Deliberation, *Theory and Decision*, 17, 71-95.

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms, *Quarterly Journal of Economics*, 75, 643-669.

Gibbard, A., & Harper, W.L. (1978). Counterfactuals and Two Kinds of Expected Utility. In P. Gärdenfors & N. Sahlin (Eds.), (1988), *Decision, Probability and Utility*, Cambridge: Cambridge University Press.

Gregory, R. (1981). *Mind in Science*, Cambridge: Cambridge University Press.

Gunderson, K. (1970). Asymmetries & Mind-Body Perplexities, M. Radner & S. Winokur eds., *Minnesota Studies in the Philosophy of Science 4*, Minneapolis: Univesity of Minnesotta Press.

Herzberger, R.C. (1970). Paradoxes of Grounding in Semantics, *Journal of Philosophy*, 67, 145-67.

Jeffrey, R.C. (1983). *The Logic of Decision*, 2nd Revised Edition, Chicago: University of Chicago Press.

Jeffrey, R.C. (2004). *Subjective Probability: The Real Thing*, Cambridge: Cambridge University Press.

Levi, I. (1975). Newcomb's Many Problems, *Theory and Decision*, 6, 161-175.

Lewis, D. (1979). Prisoner's Dilemma is a Newcomb Problem, *Philosophy & Public Affairs*, 8,3, 235-240.

Libet, B. (1985). Unconscious Cerebral Initiative & the Role of Conscious Will in Voluntary Action, *Behavioral & Brain Sciences*, 8, 529-566.

Maitzen, S. & Wilson, G. (2003). Newcomb's Hidden Regress, *Theory and Decision*, 54, 151-162.

McKay, P. (2004). Newcomb's Problem: The Causalists Get Rich, *Analysis*, 64,2, 187-89.

Nozick, R. (1969). Newcomb's Problem and Two Principles of Choice. In N. Rescher (Ed.) *Essays in Honor Of Carl G. Hempel,* Dordrecht: D. Reidel.

Nozick, R. (1993). *The Nature of Rationality*, Princeton: Princeton University Press.

Priest, G. (2002). Rational Dilemmas, *Analysis*, 62, 11-16.

Royce, J. (1900). *The World & the Individual*; reprinted 1959. New York: Dover.

Schick, F. (1979). Self-Knowledge, Uncertainty and Choice, *British Journal for the Philosophy of Science*, 30, 235-252.

Schmidt, J.H. (1998). Newcomb's Paradox Realized with Backward Causation, *British Journal for the Philosophy of Science*, 49, 67-87.

Slezak, P. (1983). Descartes' Diagonal Deduction, *British Journal for the Philosophy of Science,* 33, 41-52.

Sorensen, R.A. (1987). Anti-Expertise, Instability, and Rational Choice, *Australasian Journal of Philosophy*, 65, 3, 301-315.

Sorensen, R.A. (1988). *Blindspots*, Oxford: Clarendon.

Tversky A., & Kahneman, D. (1974). Judgement Under Uncertainty: Heuristics and Biases, *Science*, 185, 1124-1131.

Wolf, F.A. (1981). *Taking the Quantum Leap*, New York: Harper & Row.

Wright, G. (1984). *Behavioural Decision Theory: An Introduction*, Harmondsworth: Penguin.