RUSSELL DAW AND TORIN ALTER

# FREE ACTS AND ROBOT CATS

[*Philosophical Studies 102: 345-57, 2001*]

Mark Heller (1996) proposes that 'free action' is a Putnamian-kind term. In other words, Heller contends that

(H1) 'Free action' is subject to the causal theory of reference

and thus that

(H2) The essential nature of free actions can be discovered only by empirical investigation, not by conceptual analysis.

Heller's proposal, if true, would have significant philosophical implications. Consider the enduring issue we will call the Compatibility Issue (hereafter CI): whether the thesis of determinism is logically compatible with the claim that

(FR) Some human actions are free.

Nearly all attempts to settle CI have employed the following two-step process: philosophers first use conceptual analysis to ascertain the meaning of the key terms ('free', 'determinism', etc.) and then argue that, in light of their analyses, determinism is or is not compatible with FR. But if Heller's proposal is true, then the traditional method for resolving CI is fundamentally flawed. We will argue that Heller is mistaken: both H1 and H2 are false, and the traditional method for resolving CI is not based on a misguided view about the semantics of 'free action', at least not in the way Heller suggests.[1]

Views related to Heller's have been advanced for quite some time. Antony Flew (1955, 1956) and Max Black (1958) argue that the meaning of 'free action' is fixed by paradigm cases—a claim which is closely related to H1, as we will explain. In addition, a view similar to

Heller's can be derived by conjoining more recent views. Richard Double (1991, 1996) argues that 'free' is a moral term. Richard Boyd (1988) and David Brink (1988) suggest that moral terms designate natural kinds. A marriage between Double's view and the Boyd-Brink view bears surprising progeny:

(NK) 'Free action' is a natural-kind term.

On the received view about natural-kind terms, NK entails Heller's proposal. Since NK entails Heller's proposal, our arguments will indirectly challenge NK. We will also provide a direct argument against NK.[2]

We will proceed as follows. In §1, we will explain Heller's proposal in detail and show how it relates to Flew's view. We find the connection to be of more than historical interest, since—we will argue in §2—a cogent objection to Flew's view also undermines Heller's proposal. We have other objections to Heller's proposal, and we will present them in §2 as well. We will close in §3 by clarifying our view about the semantics of 'free action' and the role of scientific investigation and conceptual analysis in resolving CI.

## 1. FROM CONCEPTUAL ANALYSIS TO EMPIRICAL INVESTIGATION

Flew (1955, 1956) argues that the meaning of 'free' is fixed by paradigm cases. Peter van Inwagen summarizes Flew's view well:

There are various words and phrases we use in ascribing free action to people: besides the obvious 'acted freely' and 'did it of his own free will', there are such phrases as 'could have done otherwise', 'had a choice about what she did', 'had alternatives', and 'could have helped doing what he did'. We learn these phrases by watching people apply them in concrete situations in everyday life, just as we learn, for example, colour words. These concrete situations serve as paradigms for the application of these words: the words mean things of that sort (1983, p. 107).

On Flew's account, when we speak of free action, we mean the sort of action denoted in paradigm applications of 'free action', 'acted freely', etc.

Flew claims that if the paradigm-case theory of the meaning of 'free action' is incorrect, "then it is hard to see what meaning these expressions have and how if at all they could ever be taught, understood or correctly used" (1955, p. 151). On the other hand, he contends that if the paradigm-case theory is correct,

> then anyone who tells us that science shows or could show us that there is no such thing as acting freely, etc., is: either just wrong, because there certainly are cases such as our paradigms; or misleadingly using the key expressions in some new sense needing to be explained (1955, p. 151).

So according to Flew, given the manner in which 'free' and related phrases gain currency, no one can reasonably doubt that some actions are free.

Flew also claims that determinism and FR are compatible, and that their compatibility can be demonstrated by analyzing the meaning of 'free' and related phrases. The type of analysis he provides is conceptual: he ponders paradigm cases in order to discern the necessary and sufficient conditions for acting freely. On his analysis, the essential feature of free action is

that the agent is not compelled. Flew then argues that freedom is compatible with determinism on the grounds that a lack of compulsion is consistent with the truth of determinism (which he regards as perfect predictability).

Flew's analysis of the meaning of 'free' is basically the same as the analyses given by Hobbes (1651, 1654), Hume (1748), Moritz Schlick (1939), A. J. Ayer (1954), and numerous others. What distinguishes Flew from previous compatibilists is his claim that the meaning of 'free' is fixed by paradigms. Now suppose one held that the essence of the actions ostended in paradigm uses of 'free' can be ascertained only by empirical investigation, not conceptual analysis. That supposition brings us to Heller's proposal.

Heller claims that empirical investigation, rather than conceptual analysis, will reveal the essential nature of free actions, because he believes that Putnam's view about kind terms like 'cat' (commonly called the causal theory of reference) applies to 'free action'. According to Putnam (1970, 1973, 1975) and Kripke (1972), many general nouns serve as names for kinds—names whose extensions are determined neither by the descriptions associated with the names nor by our conception of the kinds. On Heller's reading of Putnam, "the extension is determined by paradigm cases, so that a cat is anything that is of the same kind as the paradigm cases" (1996, p. 333–4). Being the same kind is a matter of sharing a certain underlying trait (or traits), which on the Putnam-Kripke view can be discovered only by empirical investigation.[3] Thus, following Putnam, Heller states:

> To discover what it is to be a cat we put a cat before us and ask "what is essential to this thing?" We do not first have to have a concept of a cat or even know that the thing in question is called a "cat." Discovering the essential nature of the kind becomes an empirical matter (1996, p. 334).

Putnam's view has been defended as the correct account for many terms, including 'gold', 'water', 'tiger', 'lemon', and even 'unicorn'. But Heller is the first to apply the causal theory specifically to 'free action'. According to Heller, what goes for 'cats' and cats goes for 'free acts' and free acts. He states: "We have a particular action before us, and we ask 'what is essential to this kind of action' " (1996, p. 334); "what it takes to discover what is essential to free action is not conceptual analysis" (1996, p. 336); and "We must discover empirically what it is to be a free action—what it is to be one of these acts" (1996, p. 335).

Heller's proposal, if true, would have considerable import. If 'free action' is a Putnamian-kind term, then our concept of free action in no way determines the conditions for membership in the class of free actions. Just as our concept of cat might be completely misguided—cats might turn out to be robots remotely controlled from Mars (Putnam, 1970, 1975)—our concept of free action might be completely misguided. In fact, Heller argues, "It could even turn out that our concept [of free action] is self-contradictory (I suspect this is so), without this in any way implying that there is no free action or that compatibilism is false" (1996, p. 335).[4] He concludes that "general conceptual arguments and arguments by counterexample do not have the significance for the free will debate that they have been taken to have" (1996, pp. 336–7).

Although Heller (1996) seems interested only in defending compatibilism against incompatibilist counterexamples, his sword cuts both ways: if his proposal is true, then neither compatibilist nor incompatibilist analyses of the concept of free action are relevant to CI (nor are conceptual counterexamples to such analyses). If 'free action' is a Putnamian-kind term, then

whether FR is compatible with determinism will depend on what the essence of free action is, and that can be discovered only empirically.

We might empirically discover that nothing fits our concept of free action. If 'free action' is a Putnamian-kind term, then that discovery would not show that there are no free actions—any more than the discovery that nothing fits our concept of cat would show that there are no cats. However, in contrast to Flew, Heller allows that there might not be any free actions: "if the acts we take to be free form no kind at all, then 'free act' does not refer, and there are no free acts" (1996, p. 336). If 'free action' does not refer, then FR—one of the principal claims in CI—is either meaningless or false, depending on one's theory of reference-failure.

## 2. AGAINST HELLER'S PROPOSAL

Heller's article is devoted to exploring the consequences of his proposal—in particular, to arguing that his proposal undermines standard objections to compatibilism. But he provides no independent reason to accept his proposal. We will now argue that his proposal is false. For simplicity, we will grant that the causal theory applies to 'cat', 'water', 'gold' and all other terms that Putnam and Kripke regard as names for natural kinds.[5]

For our first objection, let us reconsider Heller's analogy between 'free action' and Putnamian-kind terms like 'cat', 'water', and 'gold'. 'Cat', 'water', and 'gold' are thought to refer to kinds with natural essences, and thus are regarded as natural-kind terms. The natural essence of cats is given by their DNA and historical origins; the natural essence of water is its molecular composition, $H_2O$; and the natural essence of gold is given by its atomic number, 79. Could free actions have a natural essence? We think not.

Perhaps future neuroscientists will discover that some particular type of brain state (call it B) is perfectly correlated with paradigmatically free human actions. Nevertheless, such a discovery would not be the discovery of the natural essence of free actions. Since the late 1960's, philosophers of mind have generally agreed that any acceptable theory of mental states must allow for multiple realizability—the idea that a given type of mental state can be realized in radically different physiological structures (Putnam, 1967; Block and Fodor, 1972; Boyd, 1980; LePore and Loewer, 1989; Pereboom and Kornblith, 1991). For example, it is metaphysically possible that there be creatures who experience burning pain but lack C-fibers, even though burning pain and C-fiber stimulation may be perfectly correlated in human beings. Similarly, it is metaphysically possible that there be creatures who act freely in absence of B (assuming there are free actions). Such creatures might include animals like dolphins (whose brains are structured differently from ours) or even silicon-based Martians. Theories about 'free action' and free action should not rule out such possibilities. Similar reasoning excludes any other natural property from being the essence of free actions. Thus, free actions do not have a natural essence as water, gold, and cats do.[6] (Even if water, gold, and cats do not have natural essences, their having such essences is epistemically possible.) Therefore, what goes for 'cats' and cats does not go for 'free acts' and free acts.

Our second objection concerns reference failure and the existence conditions for free acts. Consider the possibility that

(F1) The actions ostended in paradigm applications of 'free action' form no kind.

If 'free action' were a Putnamian-kind term and F1 were true, then 'free action' would fail to refer and there would be no free acts anywhere. And that conditional would remain true even if

(F2) On Mars or in Heaven there are creatures whose actions fit our concept of freedom.

Those implications of Heller's proposal are problematic. There might be creatures who act freely even if no Earthly creatures do. Thus, it would be wrong to conclude that no creatures—Martians, God, angels, etc.—could act freely, solely on the grounds that the Earthly actions called free form no kind. Yet such an inference would be warranted if 'free action' were a Putnamian-kind term. In short: Heller's proposal entails that F1 is a sufficient condition for 'free action' to fail to refer and for there to be no free actions anywhere; F1 is not a sufficient condition for that state of affairs; therefore, Heller's proposal is false.

Our third objection to Heller's proposal is based on an objection van Inwagen raises against Flew's view. Van Inwagen claims that applying the paradigm-case theory to 'free action' leads to absurdities. To demonstrate his claim, van Inwagen presents the following thought experiment:

(M) When any human being is born, the Martians implant in his brain a tiny device—one that is undetectable by any observational technique we have at our disposal, though it is not in principle undetectable—which contains a "program" for that person's entire life: whenever that person must make a decision, the device causes him to decide one way or the other according to the requirements of a table of instructions that were incorporated into the structure of the device before that person was conceived (1983, p. 109).

Suppose M turns out to be true. In that case, if the paradigm-case theory applied to 'free action', then free action would just be M-type behavior. But that, van Inwagen rightly concludes, is

absurd: if M turns out to be true, then the correct conclusion would not be that free action is M-type behavior, but rather that no human actions are free.

Van Inwagen's reductio against Flew's view works equally well against Heller's proposal. If 'free action' were a Putnamian-kind term, then the discovery that M is true would warrant the conclusion that free action is M-type behavior.[7] That conclusion is no less absurd today than it was when van Inwagen adduced it against Flew's view more than fifteen years ago. Further, consider van Inwagen's Martians, the creatures who implant the devices in the brains of human infants. Might their actions be free? If Heller's proposal were true, then their actions would be free only if they were of the same kind as the human actions described in the M-scenario. We find that absurd as well.

Heller claims that van Inwagen's criticism of Flew's view relies on the assumption that "concept determines extension" (1996, p. 336, fn. 7). Since Heller explicitly denies what he regards as van Inwagen's assumption, he might claim that his proposal is immune from van Inwagen's reductio and accuse us of begging the question. However, we are not assuming that our concept of free action determines the extension of 'free action', and thus we are not begging the question. Rather, we are appealing to intuitions about how a discovery like M is best described, and such intuitions bear directly on whether the concept of free action determines the extension of 'free action'. Indeed, appealing to such intuitions lies at the heart of the Putnam-Kripke approach. Thus, Heller cannot in fairness dismiss the relevance of intuitions about cases like M.

It is important to recognize the extent to which Heller's proposal discounts ordinary presuppositions about free action. If 'free action' were a Putnamian-kind term, then other

radically counterintuitive discoveries about free action would be possible. In §1, we touched briefly on the reasons for this. Let us now explain in more detail.

On traditional, description theories of meaning, the extension of a general noun is the set of objects which satisfy (all or a relevant subset of) the descriptions associated with the noun. In contrast, causal theorists contend that for many general nouns—those we have been calling Putnamian-kind terms—associated descriptions do not determine the noun's extension. As Steven Schwartz puts it, "At best the descriptions associated with such a term are a handy guide in picking out things of the kind named, but the descriptions do not determine what it is to be of the kind" (1977, p. 27). Thus, if the causal theory applies to 'free action', then the descriptions we associate with the term do not determine what it is for an action to be free. As Heller puts it, our concept of free action does not determine what it is for an action to be free.

Causal theorists also contend that there are no analytic claims of the sort ⌜ Every k has P ⌝ (where k is a Putnamian-kind term and P is a property) and that any claim of that sort is corrigible (Putnam, 1970). Thus, in addition to arguing that cats could turn out to be robots, Putnam (1970) argues that we could discover such things as that normal lemons are really blue and that normal tigers do not really have stripes. Kripke likewise argues that "we might . . . find out tigers had none of the properties by which we originally identified them" (1972, p. 121). Thus, if one accepts that 'free action' is a Putnamian-kind term, then one must be willing to accept that free action might be completely different from how we conceive it to be. Genuinely free action might turn out to be pre-programmed behavior (if M is true), controlled movement (if an alien neurosurgeon is manipulating our brains to cause us to decide as we do), etc.

We find such claims about free action absurd. We do not deny that M (or some such scenario) could turn out to be true. The issue, however, is not whether M is epistemically

possible, but what would follow if M were actual. We contend, with van Inwagen, that if M turns out to be true, then the correct conclusion would not be that free action is M-type behavior, as Heller's proposal implies. Rather, what we would have discovered is that the actions that we have been calling free were not really free.

For the same reason, CI would not be resolved by the discovery that M is true. CI is not, and has not traditionally been regarded as, a puzzle about whether the events ostended in paradigm uses of 'free action' are compatible with determinism. Rather, CI is the issue of whether determinism is compatible with the claim that some human actions are free—especially in a sense relevant to ascriptions of moral responsibility, and regardless of whether there are or ever have been any such actions.

For our final objection, consider the consequences of applying the causal theory to other terms in the free-will debate. If 'free action' is a Putnamian-kind term, then why not 'compelled action' as well? We see no plausible basis for Heller to object to such an extension. Yet that extension would create further trouble for his proposal. If 'free action' and 'compelled action' are both Putnamian-kind terms and each denotes a kind, then empirically discovering that there is no essential difference between actions we have been calling free and those we have been calling compelled would be the same as discovering that there is no difference between the kinds free action and compelled action themselves. That strikes us as even more absurd than the idea that free action could turn out to be M-type behavior. Any theory on which free action could turn out to be the same as compelled action is unacceptable.

We have objected to Heller's proposal on the following grounds. First, we argued that free actions do not have a natural essence and thus Heller's analogy between 'free acts' and 'cats' breaks down. Second, we argued that Heller's proposal entails the false claim that if the

actions ostended in Earthly paradigm applications of 'free action' form no kind, then there would be no free actions anywhere. Third, we argued that his proposal has the unacceptable consequence that genuinely free action could turn out to be M-type behavior, or movement produced by an alien neurosurgeon. Finally, we argued that on a natural extension of his proposal, it could turn out that there is no difference between free and compelled action. We find that consequence unacceptable as well.

## 3. CONCEPTUAL ANALYSIS REVISITED

We have argued that 'free action' should not be assimilated to Putnamian-kind terms like 'cat' and 'water'. We see no reason to believe that one should eschew conceptual analysis when attempting to resolve CI, or that conceptual counterexamples to such analyses are irrelevant to CI.

We do not deny that scientific investigation might shed light on the nature of free action. For one thing, a full elaboration of the meaning of 'free action' might involve invoking other terms to which the causal theory does apply. Further, it is not known what scientific investigation may reveal about the physiological and functional bases of actions commonly regarded as free or about action generally. Nonetheless, it is unclear how any such empirical discovery could serve as the basis for resolving CI.

We also do not deny that our concept of freedom may need revision. But there are limits. If a revised concept of freedom is to be an extension or limitation of our current concept, rather than a replacement, then the revisions must respect certain constraints. For example, a revised concept would have to exclude M-type behavior.[8]

Finally, having defended the traditional method for addressing CI, we should note that we are not suggesting that further analysis of our concept of freedom will necessarily lead to a traditional resolution of CI. Ted Honderich (1988) argues that there is no universally shared, complete concept of freedom, and Double (1991, 1996) argues that our concept of free action is inconsistent. Both arguments may be correct and may have important implications for CI.[9] In any case, our purpose in this paper has been to argue that we cannot resolve CI by the end run proposed by Heller—by dismissing our concept of freedom, however confused it may be.[10]

R. Daw
Portland, Oregon USA

T. Alter
Department of Philosophy
The University of Alabama
Tuscaloosa, Alabama 35487-0218 USA
talter@tenhoor.as.ua.edu

REFERENCES

Ayer, A. (1954) "Freedom and Necessity", in his Philosophical Essays, pp. 271–84. Macmillan, London. Reprinted in Watson, G. (ed.) Free Will, pp. 15–23, Oxford University Press, New York, 1982.

Black, M. (1958) "Making Something Happen", in Hook, S. (ed.) Determinism and Freedom in the Age of Modern Science, pp. 15–30. New York University Press, New York.

Block, N. and Fodor, J. (1972) "What Psychological States Are Not", The Philosophical Review 81: 159–81.

Boyd, R. (1980) "Materialism Without Reductionism: What Physicalism Does Not Entail", in Block, N. (ed.) Readings in Philosophy of Psychology, vol. I, pp. 67–106. Harvard University Press, Cambridge.

____. (1988) "How to Be a Moral Realist", in Sayre-McCord, G. (ed.) Essays on Moral Realism, pp. 181–228. Cornell University Press, Ithaca, New York.

Brink, D. (1988) "Legal Theory, Legal Interpretation, and Judicial Review", Philosophy and Public Affairs 17: 105–48.

Daw, R. (1995) Free Will and Determinism: An Assessment of the Traditional Approach to the Compatibility Issue, doctoral dissertation. University of Virginia, Charlottesville, USA.

Devitt, M. and Sterelny, K. (1987) Language and Reality. Basil Blackwell, Oxford.

Donnellan, K. (1983) "Kripke and Putnam on Natural Kind Terms", in Ginet, C. and Shoemaker, S. (eds.) Knowledge and Mind, pp. 84–104. Oxford University Press, New York.

Double, R. (1991) The Non-Reality of Free Will. Oxford University Press, New York.

____. (1996) Metaphilosophy and Free Will. Oxford University Press, New York.

Flew, A. (1955) “Divine Omnipotence and Human Freedom”, in Flew, A. and MacIntyre, A. (eds.) New Essays in Philosophical Theology, pp. 144–69. SCM Press, London.

___. (1956) “Philosophy and Language”, in Flew, A. (ed.) Essays in Conceptual Analysis, pp. 1–20. MacMillan, London.

Heller, M. (1996) “The Mad Scientist Meets the Robot Cats: Compatibilism, Kinds, and Counterexamples”, Philosophy and Phenomenological Research 56: 333–7.

Hobbes, T. (1651) Leviathan.

___. (1654) Of Liberty and Necessity.

Honderich, T. (1988) A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes. Clarendon, Oxford.

Hume, D. (1748) An Enquiry Concerning Human Understanding.

Jackson, F. (1998) From Metaphysics to Ethics. Clarendon, Oxford.

Kripke, S. (1972) “Naming and Necessity”, in Davidson, D. and Harman, G. (eds.) Semantics of Natural Language, pp. 253–355, plus appendices. Reidel, Dordrecht. Reprinted as Naming and Necessity, Harvard University Press, Cambridge, 1980. Page numbers in this paper refer to the reprint.

LePore, E. and Loewer, B. (1989) “More on Making Mind Matter”, Philosophical Topics 17: 175–91.

McGinn, C. (1978) “Mental States, Natural Kinds and Psychophysical Laws”, Proceedings of the Aristotelian Society 52: 195–220.

Pereboom, D. and Kornblith, H. (1991) “The Metaphysics of Irreducibility”, Philosophical Studies 63: 125–45.

Putnam, H. (1967) "Psychological Predicates", in Capitan, W. and Merrill, D. (eds.) Art, Mind, and Religion, pp. 37–48. University of Pittsburgh Press, Pittsburgh. Reprinted as "The Nature of Mental States" in Putnam's Mind, Language and Reality, pp. 429–40, Cambridge University Press, Cambridge, 1975.

____. (1970) "Is Semantics Possible?", in Kiefer, H. and Munitz, M. (eds.) Language, Belief, and Metaphysics, pp. 50–63. State University of New York Press, Albany, New York. Reprinted in Putnam's Mind, Language and Reality, pp. 139–52, Cambridge University Press, Cambridge, 1975.

____. (1973) "Meaning and Reference", The Journal of Philosophy 70: 699–711.

____. (1975) "The Meaning of 'Meaning' ", in Gunderson, K. (ed.) Language, Mind, and Knowledge, pp. 131–93. University of Minnesota Press, Minneapolis. Reprinted in Putnam's Mind, Language and Reality, pp. 215–71, Cambridge University Press, Cambridge, 1975.

Schlick, M. (1939) "When Is a Man Responsible?", in his Problems of Ethics, D. Rynin (trans.), pp. 143–58. Prentice-Hall, New York.

Schwartz, S. (1977) "Introduction", in Schwartz, S. (ed.) Naming, Necessity, and Natural Kinds, pp. 13–41. Cornell University Press, Ithaca, New York.

____. (1978) "Putnam on Artifacts", The Philosophical Review 87: 566–74.

____. (1980) "Natural Kinds and Nominal Kinds", Mind 89: 182–95.

van Inwagen, P. (1983) An Essay on Free Will. Clarendon, Oxford.

NOTES

[1]Daw (1995) formulates and rejects a proposal similar to Heller's, using some of the ideas presented in this paper.

[2]Double (1991, 1996) also rejects NK and so presumably would oppose the marriage of his view to the Boyd-Brink view. His arguments against NK are entirely different from ours.

[3]'Underlying trait' should be construed broadly so as to include facts about historical origins. See Kripke (1972).

[4]Double (1991) draws the conclusion that free actions do not form a kind (natural or otherwise) on the grounds that we make use of conflicting conceptual paradigms of free action. Heller's proposal, if true, would undermine Double's argument, since Heller's proposal implies that the extension of 'free action' is not determined by conceptual paradigms of free action.

[5]It is important to recognize that one may, without inconsistency, endorse the causal theory for some types of referring terms but not others. See Donnellan (1983), Devitt and Sterelny (1987, Chapter 5), and Schwartz (1977, 1978, 1980).

[6]See McGinn (1978) for a parallel argument regarding mental states. McGinn appeals to multiple realizability to argue that mental states are not natural kinds and thus that mental predicates are not natural-kind terms.

[7]Whenever we suppose M is true, we assume that M-type behavior forms a kind, just as Putnam (1970) seems to assume that robot cats form a kind when he imagines cats turning out to be robots.

[8]For a recent discussion of the role of conceptual analysis in metaphysics and ethics, see Jackson (1998). See pages 44–5 for his discussion of the concept of freedom.

[9]Honderich (1988) concludes that both compatibilism and incompatibilism are false; Double (1991, 1996) argues for the "non-reality of free will".

[10]For their helpful suggestions, we would like to thank Peter Dlugos, Ned Hall, Amy Karofsky, Dana Nelkin, James Otteson, Stuart Rachels, Norvin Richards, Samuel Rickless, George Thomas, an anonymous Philosophical Studies referee, and those who participated in the discussion following a presentation of this paper to the Department of Philosophy at Florida State University.