
NATURALISM AND THE PARADOX OF REVISABILITY

BY

MARK COLYVAN

Abstract: This paper examines the paradox of revisability. This paradox was proposed by Jerrold Katz as a problem for Quinean naturalised epistemology. Katz employs diagonalisation to demonstrate what he takes to be an inconsistency in the constitutive principles of Quine's epistemology. Specifically, the problem seems to rest with the principle of universal revisability which states that no statement is immune to revision. In this paper it is argued that although there is something odd about employing universal revisability to revise itself, there is nothing paradoxical about this. At least, there is no paradox along the lines suggested by Katz.

It is, to say the least, embarrassing when one's own theory turns out to be self refuting. Indeed, there is no more damning critique of a position than to show that it is either inconsistent or otherwise fails *by its own lights*. Such was the plight of naïve set theory, naïve truth theory, logical positivism and, arguably, relativism about truth. The problem in each case is that, under diagonalisation, the theory in question leads to paradox.

Jerrold Katz (2000) argues that there are similar problems for Quinean naturalism. According to Katz, the three central tenets of Quinean naturalised epistemology taken together are inconsistent. In the next section I examine Katz's argument. In section 2 I propose a solution to Katz's puzzle. In the following section I discuss a more formal argument to paradox, and in the final section I suggest that Katz's argument is self undermining.

1. *The puzzle*

In his book *Realistic Rationalism* (Katz, 2000, pp. 72–74) Katz suggests that three theses lie at the heart of Quinean naturalised epistemology. These theses are:

Pacific Philosophical Quarterly 87 (2006) 1–11

© 2006 The Author

Journal compilation © 2006 University of Southern California and Blackwell Publishing Ltd.

- (N) the principle of noncontradiction,
- (R) the principle of universal revisability,
- (S) the principle of simplicity.

The first tells us that when we find a system of beliefs to be contradictory, we must revise them. The second principle tells us that we may revise any of the beliefs in the system so as to restore consistency — no belief is immune from revision. The third, the principle of simplicity, tells us to restore consistency in the way that does least violence to the system as a whole.¹ These principles, Katz points out (2000, p. 73), are constitutive principles of Quine's naturalised epistemology. That is, they provide the mechanism for revising beliefs.

But now here comes the puzzle:

Since the constitutive principles [N, R, and S] are premises of every argument for belief revision, it is impossible for an argument for belief revision to revise any of them because revising any one of them saws off the limb on which the argument rests. Any argument for changing the truth value of one of the constitutive principles must have a conclusion that contradicts a premise of the argument, and hence must be an unsound argument for revising the constitutive principle (Katz, 2000, p. 73).

As an illustration of the problem, Katz considers an argument for revising the law of noncontradiction:

Given universal revisability, the principle of non-contradiction is revisable in principle. If it is revisable in principle, there is a possible belief-revision argument for its reevaluation. But, as we have seen, since the principle of noncontradiction is a constitutive principle, it must appear as a premise of the argument. But if it is right to revise a belief in the system, that belief was wrong all along, and if it was wrong all along, it cannot be part of a sound argument. The argument for revising the belief would be unsound and provide no grounds for the revision. Hence, there can be no sound argument for revising the principle of non-contradiction, and it is not open to revision. Nonetheless, since all beliefs are revisable, the principle of noncontradiction must be revisable, and hence it is both revisable and not revisable (Katz, 2000, pp. 73–74).

It is worth noting that it is really universal revisability that's doing the work here. Indeed, the application of universal revisability to itself would seem to yield a more direct route to paradox. If this is right, then Katz might have cast his net wider: it's not just a problem for Quinean naturalism, it's a problem for any epistemology that subscribes to a principle like universal revisability.² This, of course, does not mean that the revisability paradox is not a problem for Quinean naturalism. After all, pointing out that your favourite theory isn't alone in being self-refuting is little comfort.³ A more direct defence of Quinean naturalism is thus required. Although there have been a couple of critiques of Katz's argument,⁴ these

critiques fail to recognise a crucial point: Katz's argument simply does not generate a paradox.

2. *The solution*

As I understand Katz, he is suggesting that the problem is that we have some belief set, B , along with a set of constitutive principles,⁵ $C = \{N, R, S\}$, and from this we try to mount an argument for rejecting one of the elements of C — N , say. That is, we wish to show that

$$(1) \quad B \cup C \vDash \neg N.$$

But because $N \in C$ it would seem that we have a problem.

Katz's first shot at articulating the problem is to suggest that any argument we mount for $\neg N$ will be unsound since it will rely on the false premise N (2000, p. 73). But this is way too strong, for this would rule out *reductio ad absurdum* as a legitimate proof procedure. (One way of looking at a *reductio* proof is as a case of proving P from the (discharged) assumption of $\neg P$.)⁶ After all, if we claim that P entails $\neg P$ there's no paradox, that just tells us that P is false. Now return to argument (1). The structure is that of a *reductio* proof — there is simply nothing paradoxical about it. It's just a good argument for rejecting one of the elements of $B \cup C$; which one we reject, it would seem, is up to us. We are free to reject N if we please. I see nothing paradoxical about that, at least in the abstract.

Perhaps stating it in the abstract, though, is the problem. Perhaps what Katz really means is that one cannot use a premise that's a *constitutive principle* of the methodology to mount an argument for revising the principle in question.⁷ But then the problem is more like using a *reductio* proof to raise problems for the methodology of *reductio* proofs, or, more generally, for classical logic. If that is Katz's point, then we need to proceed very carefully. For despite this way of putting the problem, Katz seems to be assuming an underlying deductive system. But which deductive system are we talking about? That is, how are we to understand ' \vDash ' in (1)? Surely this is not classical consequence, since we're contemplating rejecting the law of non-contradiction. And since we're explicitly discussing belief revision, presumably we're operating in some non-monotonic logic appropriate for this task.

This, however, might seem a bit quick. It might be argued that we cannot simply move to a non-monotonic logic because, after all, Quine too argued that universal revisability might be invoked to overthrow any logical principle. And it might be maintained that Quine did so from within the framework of classical logic. We might see Katz as simply following Quine on this issue. While it's true that Quine held that classical first-order logic was the one true logic, the Quine of "Two Dogmas of

Empiricism" (Quine, 1980) at least, entertained the idea of alternate logics. Moreover, in "Two Dogmas" Quine held that logical theory change might result from empirical evidence. So I take it that the concern raised here really amounts to the question of how could we argue for a non-classical logic while starting out (at least tentatively) believing in classical logic? Or alternatively, mightn't there be some subset of classical principles (or close relatives) that give rise to some close relative of classical logical consequence? That is, we might not need to resort to a non-monotonic logic.⁸

The answers to the questions raised in the last paragraph are long and involved.⁹ But let me say a little by way of appeasing the worry here. Let's consider some of the alternatives for ways we might argue about fundamental logical principles. One way such a debate might go is for both parties to (tentatively) accept whatever undisputed logical principles there might be. So, for example, in a debate between paraconsistent and classical logicians, both parties might agree to (tentatively) accept *modus ponens*, but reject the classical principle of disjunctive syllogism. How far might this take the debate? Well, we might get lucky and prove that from this impoverished set of logical principles one or other of the two logics under consideration is inadequate. Though, when the debate concerns such apparently fundamental logical principles as the law of non-contradiction, there will not be a great deal of common ground on which the two protagonists can stand.¹⁰ But, even if the debate can proceed with the impoverished set of principles, eventually we wish to settle on the details of a logic. We cannot start out with a notion of logical consequence (or impoverished set of inference rules) that remains fixed throughout the debate. We must hold open the possibility of revising the notion of logical consequence as we go. If this is right, we will need to look further afield for how such a debate over logical principles might proceed.

Another way this debate might go is to accept (again tentatively) classical logic and try to show that it is inadequate *on its own terms*. Indeed, if one were to do this, it would provide a very powerful argument for rejecting classical logic. This would provide a *reductio* of classical logic and it would be clear that classical logic must be rejected (as it was clear that the conjunction of classical logic and naïve set theory had to go). Now it's true that nothing in such a demonstration of the untenability of classical logic would indicate which logic we should adopt in its place. But an examination of where the problems arose would give us some idea of what the revisions would need to be. Indeed, paraconsistent logics reject disjunctive syllogism for precisely this reason: we start with classical logic, we notice that it's explosive,¹¹ we then pinpoint the culprit — disjunctive syllogism — and we reject it. The result is a paraconsistent logic. Of course this does not yield a unique such logic, but let's suppose for the moment that it does. Let's suppose that we settle on some paraconsistent logic (RM₃, say), then we wish to continue the debate over which logical principles we

ought to accept. Since the logic has changed, we now have a different consequence relation; it's now the RM_3 turnstyle (or so we are supposing). How do we go about representing this change of turnstyle mid-debate? One way is to represent the whole debate in a belief-revision logic where the logical principles are taken to be part of the belief sets. Another way is to reject the formal framework altogether. As I suggested earlier, because debates about logic must, it would seem, presuppose some logic, they are very difficult to represent formally. It might be that there is no fixed, formal notion of logical consequence that is adequate for the task. Some kind of informal reflective equilibrium approach may well be the only way to go.¹²

So as far as I can see, we have three alternatives: (i) either we presuppose some fragment of classical logic on which all parties agree, (ii) we change logical principles (and consequence relations) along the way, (iii) we abandon formal frameworks altogether and invoke some (informal) reflective equilibrium approach. Notice that in all three cases the consequence relation is not classical (although in case (ii) it might start out that way). I've suggested that the best way to represent case (ii) is by invoking a non-monotonic belief revision logic and, indeed, no harm is done if we also invoke such a logic in case (i). Indeed, for current purposes, case (i) is not different in any relevant respects from case (ii). It's just that in case (i) we're assuming that we are not revising our impoverished set of logical principles, but surely, given the context, we need to entertain the possibility of adding, and possibly later abandoning, other logical principles. So assuming, as I have, that we are conducting the debate in a non-monotonic belief revision logic is certainly not out of order for cases (i) and (ii). What of case (iii)? If there is no formal notion of logical consequence at all then Katz's argument simply fails, because his argument clearly requires some notion of logical consequence.¹³ So in assuming that there is one fixed notion of logical consequence, albeit a non-monotonic one, I am in fact conceding a great deal to Katz. So, whichever way we look at it, we cannot think of the turnstyle in (1) as classical consequence. Indeed, given the dynamics of belief revision about logic, it is hard to think of the turnstyle in question as representing any one fixed notion of logical consequence. Still, for the sake of argument, it is helpful to think in these terms and, I contend, that some non-monotonic consequence relation is the best candidate (though I admit that I've not said enough here to render this claim uncontroversial).

How is all this relevant to Katz's argument? Well, let's suppose we're operating in some belief-revision logic. Then the truth or falsity of N is beside the point. What's important is whether N is rationally defensible. Now let's suppose that at time t_1 , prior to making inference (1), N was rationally defensible, but at some later time t_2 , after making the inference (1), N was not rationally defensible. Katz's challenge, I take it, is to provide a rational

defence of $\neg N$ at time t_2 . But isn't the answer here straightforward? As a result of (1), we rejected N and replaced it in our belief set with $\neg N$. So we have $\neg N \in B^* \cup C^*$ (where B^* and C^* are the new, revised belief sets and constitutive principles respectively). So it's trivial that¹⁴

$$(2) \quad B^* \cup C^* \models \neg N.$$

So we do have an argument for $\neg N$. It's not the one we originally employed (i.e., (1)); instead, we find ourselves with a new argument (i.e., (2)). We might say we saw off the limb on which we originally rested, but we always find ourselves resting on a new limb. In effect, Katz assumes that the original argument for revising a constitutive principle is the only one ever available. But this is clearly mistaken. Indeed, his mistake seems to result from a failure to take into account the dynamic nature of the inferences here.¹⁵

Katz illustrates the revisability paradox with another example — revising R itself — and in this example a slightly different concern emerges. So let's consider revising R . Here we have an argument

$$(3) \quad B \cup C \models \neg R.$$

This just tells us that our premise set is inconsistent — no paradox so far. But what do we do about the inconsistency? We can revise any of the premises in the argument, but let's suppose we choose to revise R to R^* (where the new principle R^* might be some restricted version of R).¹⁶ How is this paradoxical? We can't (easily) argue from our new set of principles for the revision of R^* back to R , say, but no-one claimed we could. I don't see anything paradoxical about this. Indeed, this is no more puzzling than democratically voting for a dictatorship. You can't go back again (from a dictatorship to a democracy — at least not without a revolution), and there's no doubt something a bit odd about voting for a dictator, but there's nothing paradoxical about it. So too with the 'paradox' of revisability. To revise your revisability principle is an odd thing to do (since you can't easily retrieve the original revisability principle), but there's no paradox here.¹⁷

3. *A proof of the paradox?*

Katz has one final shot at formulating his misgivings about universal revisability. He claims that right from the outset we should have been wary of this principle:

Looked at from the right angle, universal revisability already flashes the signal Paradox! Paradox! Paradox!. Unrestricted universality sanctions the dangerous move of self application,

which is a familiar feature of paradox. From the application of the belief-revision epistemology to itself, it follows that a revisability principle is unrevisable. Hence, just as the barber paradox proves that there is no actual barber who shaves all and only those who do not shave themselves, the revisability paradox proves that there is no actual epistemology that says that everything including itself is revisable (Katz, 2000, p. 74).

The first thing to note is that the comparison with the barber paradox is not appropriate. In the barber paradox, even before the derivation of the paradox, we have absolutely no reason to suppose that there is such a barber. The derivation of paradox just seals the deal. The case of the revisability paradox is quite different. Even if there were a genuine paradox here, the case would be more like Russell's paradox, where we have a paradox emerging from an independently-plausible system. As we well know, one must proceed with due caution in such cases.¹⁸ Second, it would seem that Katz's "paradox alarm" is a little too sensitive. We all know that self reference is not a sufficient condition for paradox, and some argue that it is neither necessary nor sufficient.¹⁹

But putting such quibbles aside, perhaps we can do better than Katz's general concerns about self reference here and produce a formal argument for the paradox. To this end, I offer the following.²⁰ In any case, this argument seems to cut to the heart of the matter.

Let's formalise universal revisability (R) as $(\forall x)\diamond Rx$, where $Ra = 'a$ is revised', then the argument to contradiction is as follows:²¹

1. $(\forall x)\diamond Rx \quad w_1$ (universal revisability).

But if we apply 1 to itself and read 'is revised' as accepting its negation, then we get:²²

2. $\diamond\neg(\forall x)\diamond Rx \quad w_1$ (instantiating for 1 in 1);
3. $(\exists x)\Box\neg Rx \quad w_2$ (from 2);
4. $\Box\neg Ra \quad w_2$ (instantiating for a in 3);
5. $\diamond Ra \quad w_1$ (instantiating for a in 1);
6. $Ra \quad w_3$ (from 5).

We can derive the contradiction:

7. $\neg Ra \quad w_3$ (from 4),

so long as w_2 has access to w_3 . But this will only occur if the access relation is transitive and symmetric — in logics such as S5 (D5 or K5 would do just as well).

So, on pain of contradiction, anyone who subscribes to universal revisability cannot adopt a symmetric, transitive modal logic (like S5, for instance) as their epistemic logic. This, however, should be no surprise, for

there are well-known and compelling arguments as to why the appropriate modal system for epistemology cannot be as strong as S5. The appropriate logic is usually thought to be a non-symmetric logic somewhere in the vicinity of S4 (or D4).²³ In short, this more formal modal argument fails because the contradiction can only be derived if we adopt a fairly strong modal system — one that's inappropriate as an epistemic logic.²⁴

4. *A parting (cheap) shot*

It seems that no paradox can be generated for Quinean epistemology without invoking a logic that's inappropriate for the task at hand. Moreover (and here comes the cheap shot), it might be argued that it's not Quinean epistemology that is paradoxical but, rather, it's Katz's argument that's self refuting. Katz's argument is supposed to show that there can be no good argument for rejecting the constitutive principles (N, R, and S). That is, Katz argues that Quinean epistemology (that suggests, via R, that we can, at least in principle, revise N, R, and S) is indefensible and we ought to abandon it. If Katz's argument were to succeed, it would have delivered precisely what he argues is impossible — an argument for rejecting N, R, and S. So Katz's argument succeeds only if it doesn't succeed.

I've argued that the so-called paradox of revisability is not paradoxical at all. The problem with Katz's argument is that it does not take into account the dynamics of belief revision. In particular, it seems that the only arguments to paradox that can be plausibly defended are arguments that require logics that are inappropriate as logics of belief revision (classical logic or the modal system S5). If we restrict our attention to the appropriate logics here — for example, non-monotonic belief revision logics — we find that we can indeed mount arguments for revising the constitutive principles in Quine's naturalised epistemology. There is undoubtedly something odd about revising one's revisability principle, but there is nothing paradoxical about it. At least, there is no argument to paradox to be found in Katz's writing on this topic. Similarly, democratically voting for a dictator may well be unwise, but it is not paradoxical.²⁵

Philosophy Program
University of Queensland

NOTES

¹ Quine calls the application of simplicity in belief revision "the maxim of minimum mutilation" (Quine, 1992, pp. 14–15). See also Quine (1969) for a classic statement of naturalised epistemology.

² Indeed, Katz says as much at one stage: “there is no actual epistemology that says that everything including itself is revisable” (Katz, 2000, p. 74). In light of this, Quine’s own views on logic, for instance, are not terribly relevant here. For although Katz clearly has Quine in his sights as the main target, my aim is to show that Katz is wrong when he says that there is no such epistemology (though I hope to do this without departing too far from the basic Quinean epistemology).

³ Imagine the logical positivists mounting a defence of positivism by appeal to the inconsistency of naïve set theory.

⁴ See Resnik and Orlandi (2003) and Adler (2003). Resnik and Orlandi (2003) entertain a solution that bears some similarity to what I develop below, but they eventually settle on restricting the scope of the principle of universal revisability so that this principle itself does not fall within its own scope. This strikes me as *ad hoc* and conceding too much to Katz.

⁵ Strictly speaking, the constitutive principles will be members of the belief set as well, but it will be more convenient to keep the two separate. So, if you like, think of *B* as all the non-constitutive beliefs.

⁶ Of course, there are those who do reject *reductio* proofs, but defending intuitionism isn’t Katz’s purpose here.

⁷ Katz suggests as much in a couple of places. In particular, he says: “It might make sense to contemplate replacing any and every plank in Neurath’s boat, but it makes no sense to contemplate replacing the basic principle of ship construction that says that there have to be planks between us and the water” (Katz, 2000, p. 74).

⁸ I thank an anonymous referee of this journal for raising this issue.

⁹ See Bueno and Colyvan (2004) for some of this long and involved story.

¹⁰ This is why David Lewis (1998, p. 101) suggests that debates with those who reject the law of non-contradiction are impossible: “They have called so much into question that I have no foothold on undisputed ground.” While I don’t share Lewis’s pessimism (Bueno and Colyvan, 2004), I do agree that there will be very little common ground.

¹¹ Everything follows from a contradiction in classical logic.

¹² Indeed, this is what I take to be the case (Bueno and Colyvan, 2004).

¹³ It might be argued that in this case Quine’s epistemology is also in trouble, since it too relies on an underlying formal notion of logical consequence. I don’t agree with this assessment, but I won’t go into details here. We are, after all, interested in Katz’s objections to Quine’s epistemology and this new objection (if it is one) is certainly not what Katz had in mind.

¹⁴ It might be complained that this argument is question begging, but that’s not Katz’s concern. He claims that there is *no* argument for $\neg N$. In any case, the circularity is simply the circularity of proving an axiom from itself or justifying an axiom by its consequences. This may be the best we can do when asked to provide justification for fundamental principles or axioms.

¹⁵ See Bueno and Colyvan (2004) for replies to other, related objections to revising the law of non-contradiction.

¹⁶ As I already mentioned, Resnik and Orlandi (2003) suggest restricting the scope of *R*, in the light of Katz’s paradox.

¹⁷ Similarly, fallibilists generally allow that their doctrine of fallibilism might be mistaken and that we might in fact be infallible. Again there’s no contradiction in this.

¹⁸ Although perhaps Katz’s point might be strengthened by pointing out that the Barber paradox proves more than that there is contingently no such barber, it proves that such a barber is *impossible*. Similarly, the revisability paradox might be thought to show the impossibility of Quinean epistemology.

¹⁹ Yablo’s paradox is supposed to be an example of a liar-like paradox that is non circular. Stephen Yablo (1993) and Roy Sorensen (1998) have argued that self reference (or circularity)

are neither necessary nor sufficient for paradox. Others such as Priest (1997) remain unconvinced.

²⁰ I am indebted to my colleagues Dominic Hyde, Gary Malinas and Calvin Normore for their help with the following formal argument. I should add that this is the only such argument to paradox that we could produce from universal revisability alone.

²¹ Of course Quine, who rejects modal notions, will be unhappy with the following presentation of the argument. Be that as it may, "revisability" is clearly a modal notion and is naturally formalised in the way I've suggested.

²² As Mike Resnik pointed out to me, this step is rather contentious. Taking *Ra* to be equivalent to accepting $\neg a$ is one thing, but it's another thing entirely to assume that *accepting* $\neg a$ is equivalent to $\neg a$. After all, it certainly does not follow that accepting some proposition implies that the proposition is true. But we might think of the whole proof as being about acceptance rather than truth. That is, we might think of the whole proof as being inside the scope of a "rationally acceptable" operator. In any case, I will overlook this problem because there is another, more serious problem on the way.

²³ See Girle (2000) and Hintikka (1963).

²⁴ But as James Chase pointed out to me, there is a quicker route to paradox along the lines presented above. Instead of instantiating for 1 in line 1, instantiate for some logical truth. This leads immediately to paradox in any normal modal logic. Although this is clearly not the paradox Katz had in mind, it does suggest that there is something fundamentally wrong with the formal presentation. In fact, it would seem that Resnik's concerns about treating '*a* is revised' as equivalent to ' $\neg a$ ' (along with an unrestricted domain of quantification) is more contentious than I suggested earlier. Moreover, it may be that the revisability operator involves both alethic and epistemic modalities (again thanks to James Chase). Although I won't pursue such issues here, it would be interesting to investigate other ways of formalising Katz's argument.

²⁵ An earlier version of this paper was presented to the Department of Philosophy at the University of Queensland and at the 2003 Australasian Association of Philosophy Conference at the University of Adelaide. I'd like to thank the audiences on both occasions for many helpful criticisms and suggestions. I am also very grateful to Carol Booth, Otávio Bueno, James Chase, Jennifer Fisher, Dominic Hyde, Gary Malinas, Daniel Nolan, Calvin Normore, Graham Priest, Jack Smart, and Juliana Weingaertner for helpful discussions or comments on an earlier draft. I am particularly indebted to Mike Resnik for an initial conversation and ongoing correspondence on Katz's puzzle. I'd also like to thank an anonymous referee of this journal for several helpful suggestions. Work on this paper was funded by the Australian Research Council (grant number DP0209896).

REFERENCES

- Adler, J. (2003). "The Revisability Paradox," *Philosophical Forum* XXXIV, pp. 383–389.
- Bueno, O. and Colyvan, M. (2004). "Logical Non-Apriorism and the Law of Non-Contradiction," in G. Priest, J. C. Beall, and B. Armour-Garb (eds.), *The Law of Non-Contradiction: New Philosophical Essays*. Oxford: Oxford University Press, pp. 156–175.
- Girle, R. (2000). *Modal Logics and Philosophy*. Chesham: Acumen.
- Hintikka, J. J. (1963). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, NY: Cornell University Press.
- Katz, J. J. (2000). *Realistic Rationalism*. Cambridge, MA.: MIT Press.
- Lewis, D. K. (1998). "Logic for Equivocators," in *Papers in Philosophical Logic*. Cambridge: Cambridge University Press, pp. 97–110.
- Priest, G. (1997). "Yablo's Paradox," *Analysis* 57, pp. 236–242.

- Quine, W. V. (1969). "Epistemology Naturalized," in *Ontological Relativity and Other Essays*. New York: Columbia University Press, pp. 69–90.
- Quine, W. V. (1980). "Two Dogmas of Empiricism," in *From a Logical Point of View*, second edition. Cambridge, MA.: Harvard University Press, pp. 20–46.
- Quine, W. V. (1992). *Pursuit of Truth*, revised edition. Cambridge, MA.: Harvard University Press.
- Resnik, M. D. and Orlandi, N. (2003). "Holistic Realism: A Response to Katz on Holism and Intuition," *Philosophical Forum* XXXIV, pp. 301–315.
- Sorensen, R. A. (1998). "Yablo's Paradox and Kindred Infinite Liars," *Mind* 107, pp. 137–155.
- Yablo, S. (1993). "Paradox without Self-Reference," *Analysis* 53, pp. 251–252.