STEPHEN MAITZEN

THE KNOWER PARADOX AND EPISTEMIC CLOSURE*

ABSTRACT. The Knower Paradox has had a brief but eventful history, and principles of epistemic closure (which say that a subject automatically knows any proposition she knows to be materially implied, or logically entailed, by a proposition she already knows) have been the subject of tremendous debate in epistemic logic and epistemology more generally, especially because the fate of standard arguments for and against skepticism seems to turn on the fate of closure. As far as I can tell, however, no one working in either area has emphasized the result I emphasize in this paper: the Knower Paradox just *falsifies* even the most widely accepted general principles of epistemic closure. After establishing that result, I discuss five of its more important consequences.

1. INTRODUCTION

This paper tries to bring together two things that ought to have met but seem not to have made each other's acquaintance: (a) the Paradox of the Knower, the focus of a small number of impressive technical discussions,[1] and (b) the debate over epistemic closure, the subject of an enormous, usually less technical, literature.[2]

The Knower Paradox, like the much older Liar Paradox, uses seemingly correct inferences to derive an unwelcome conclusion from seemingly correct premises.[3] Principles of epistemic closure say that knowledge is closed under such operations as known material implication and known logical entailment – they say that a subject automatically knows any proposition she knows to be materially implied, or logically entailed, by a proposition she already knows. (I distinguish such principles from principles of *doxastic* closure, which say that belief or justified belief is closed under similar operations.)

The Knower Paradox has had a brief but eventful history, and epistemic closure has been the subject of tremendous debate in epistemic logic and epistemology more generally, especially because the fate of standard arguments for and against skepticism seems to turn on the fate of closure.[4] As far as I can tell, however, no one working on either topic has emphasized the fact I emphasize in this paper: the Knower Paradox just *falsifies* even

the most widely accepted general principles of epistemic closure. After establishing that fact, I will discuss five of its more important consequences.

2. THE KNOWER PARADOX

One version of the (Strengthened) Knower Paradox[5] starts with the following plausible assumptions:

(A1)   (A sentence S is true) $\equiv$ (S expresses a true proposition)

(A2)   (A sentence S is false) $\equiv$ (S expresses a false proposition)

(A3)   (A sentence S is known to be true) $\supset$ (S expresses a proposition which is known to be true)

(A4)   (A proposition P is known to be true) $\supset$ (P is true)

These assumptions seem innocent enough, although later on I will consider some reasons one might have for rejecting the first three assumptions, reasons which, in the end, take none of the sting out of the paradox. The fourth assumption is just a conceptual truth about knowledge. An immediate corollary of (A3) is

(C)   If a sentence S fails to express a proposition, then the proposition that S is not known to be true is a *true* proposition.

The Knower Paradox then uses a self-referential sentence like that used by its close cousin, the Liar,

(G)   It is not the case that sentence G is known to be true,

and reasons as follows:

(1)   (G is false) $\supset$ (G is true).

(2)   ((G is false) $\supset$ (G is true)) $\supset$ (G is true).

(3)   Therefore, G is true.

I will say more about premise (1) in a moment. Premise (2) is just the rule of *consequentia mirabilis*, itself just a version of *reductio ad absurdum.* The conclusion, (3), follows from (1) and (2) by *modus ponens.* We can

assure ourselves that the argument from (1) and (2) to (3) is valid by noting that (1) and (2) are logically equivalent, respectively, to

(1*)   ∼(G is false) ∨ (G is true)

and

(2*)   ∼(1*) ∨ (G is true),

from which (3) follows by disjunctive syllogism.

Premise (1) gets justified this way: (i) If G is false, then G *is* known to be true; (ii) by A3, then, G expresses a proposition which is known to be true; (iii) by A4, any proposition which is known to be true is true; (iv) by A1, then, G is true. Hence, if G is false, then G is true.

Thus, we seem to have a straightforward demonstration of the truth of sentence G, a pretty mundane result until we look again at G itself. According to A3, G is known to be true only if G expresses a proposition which is known to be true. But if G expresses a proposition, that proposition is *not* known to be true. Here's why: (i) if G expresses a proposition, then that proposition is either true or else not true; (ii) if G expresses a true proposition, then G expresses a proposition which is not known to be true; (iii) if, on the other hand, G expresses a proposition which is not true, then, by A3 and A4, G is not known to be true; (iv) either way, then, G is not known to be true.

So, in spite of our having *demonstrated* that G is true, it is impossible that G should be known to be true. For, by A3 and A4, if G is known to be true, then G expresses a true proposition; but if G expresses a true proposition, then G expresses a proposition which is not known to be true, in which case, by A3 again, G is not known to be true. Thus, if G is known to be true, G is not known to be true; by *reductio*, then, G is not known to be true – a result we have proved by reasoning that works in any possible world. So we seem to have demonstrated the truth of a sentence whose truth it is demonstrably impossible to know. In spite, then, of having demonstrated that a particular conclusion is true, we are logically debarred from knowing that it is true. "What good is demonstration, then?" one might ask. Welcome to the Knower Paradox.

3. EPISTEMIC CLOSURE

For my purposes, the important consequence of the Knower Paradox is the trouble it makes for standard principles of epistemic closure.[6] These principles have, of course, caused controversy at least since Jaakko Hintikka

championed a particularly strong version of closure in his epistemic logic some thirty-five years ago (Hintikka 1962); Fred Dretske's rejection of even weaker closure principles six, eight, and nine years later only heightened the controversy (Dretske 1968; 1970; 1971). If the paradox resists dissolution, and if I am right about its consequences, then the case against closure can finally be closed: in their most general forms, the standard principles of epistemic closure are all false.

Consider, first, the best-known of all the principles of epistemic closure, the principle of closure under known implication:[7]

(E1) $\Box \, \forall s \, \forall \phi \, \forall \psi \, ((Ks\phi \, \& \, Ks(\phi \supset \psi)) \supset Ks\psi),$

where the variable s ranges over epistemic subjects and ϕ and ψ are arbitrary objects of knowledge (sentences or propositions, depending on your preference). According to E1, anyone who knows that ϕ and knows that ϕ implies ψ automatically knows that ψ. This principle has struck some philosophers as obviously true, while striking others as obviously false. But the principle's staying power in the face of repeated attacks testifies both to its resilience and to the failure of any of those attacks to count as a knock-down refutation.

Some proposed counterexamples to E1 involve a subject who knows that ϕ and that ϕ implies ψ but who fails to "connect" those objects of knowledge well enough to come to know that ψ.[8] The problem with such counterexamples is that the defender of closure can reply, not implausibly, that connecting ϕ and ψ in the appropriate way is a precondition for knowing that ϕ implies ψ; thus, she can say, the subject in question doesn't really know that the implication holds.[9] Or she can reply by hedging a bit, saying that E1 holds for certain, perhaps idealized, epistemic subjects – those who infallibly connect ϕ and ψ, in the way required by E1, whenever they know that ϕ implies ψ.

Against the counterexample supplied by the Knower Paradox, however, neither of those replies will work. Let s consider our paradoxical reasoning as long as it takes to appreciate the validity of that reasoning; it shouldn't take long for her to come to know that our three-step argument is deductively valid – or, equivalently, for her to come to know that the conjunction of its two premises implies its conclusion: $Ks[((1) \, \& \, (2)) \supset (3)]$. After a bit more reflection she'll also come to know the conjunction of those two premises. We've already seen why G cannot be false without being true: recall the reasoning that established premise (1). Premise (2) is an instance of an ancient logical truth; if it can't be known to be true then neither *consequentia mirabilis* nor *reductio ad absurdum* nor disjunctive syllogism can be known to be valid forms of inference. Premises (1) and

(2) are, moreover, entirely compatible with each other: there is no reason they cannot both be true, and – except for the reason allegedly supplied by E1 – there is no reason their conjunction cannot be known by *s* to be true. But, whoever *s* is and however long she reflects, she cannot know the truth of (3), the conclusion she knows to be implied by the conjunction of (1) and (2). So principle E1 fails, however idealized the epistemic subject and however good she is at connecting the objects of her knowledge.

In the face of counterexamples, defenders of epistemic closure have retreated to logically weaker principles, such as the closure of knowledge under known implication plus belief in the consequent:[10]

(E2) $\Box\ \forall s\ \forall\phi\ \forall\psi\ ((Ks\phi\ \&\ Ks(\phi \supset \psi)\ \&\ Bs\psi) \supset Ks\psi)$.

Unlike E1, principle E2 explicitly requires that *s* believe the consequent which she knows to follow from the antecedent she already knows (and thus, presumably, already believes). While E2 survives the sorts of counterexamples it was designed to survive (Hales 1995, 192), it proves utterly ineffective against the Knower Paradox, since the conclusion, (3), of our paradoxical reasoning is impossible for any subject *s* to know even if *s* believes that (3). It seems perfectly possible, moreover, for a subject *s* to believe step (3): A1 and A2 entail that, whether or not G expresses a proposition, (3) cannot help expressing a proposition, and the only question is whether (3) expresses a true one or a false one, i.e., whether (3) is true or false. But (3) may well be true, and someone convinced by our paradoxical reasoning may well (indeed, perhaps ought to) believe that (3) is true. In any case, however, no one can possibly know that (3) is true.

What about a still weaker principle, such as

(E3) $\Box\ \forall s\ \forall\phi\ \forall\psi\ (Ks\phi\ \&\ Ks(\phi \supset \psi)\ \&\ Bs\psi\ \&\ (Bs\psi$ based on $(Bs\phi\ \&\ Bs(\phi \supset \psi))) \supset Ks\psi)$,

the closure of knowledge under known implication plus belief in the consequent based on deduction via *modus ponens*? (Hales 1995, 193). Again, no dice. Principle E3 also survives just those counterexamples to E2 that E3 was designed to survive; indeed, one recent author who regards every other standard principle of epistemic or doxastic closure as false regards E3 as at least trivially true (ibid.). For all that, though, E3 is false; it still falls victim to the Knower Paradox. Let *s* come to believe conclusion (3) by using *modus ponens* to deduce – and thereby come to believe – (3) from her knowledge of the conjunction of (1) and (2) and her knowledge of the implication between that conjunction and (3). In fact, that's probably how an ordinary person *would* come to believe (3); only an extraordinary

person would come to believe (3) spontaneously and non-inferentially. In other words, let the antecedent inside the universal scope of E3 be satisfied as completely as you like; where "ψ" stands for step (3) of our paradoxical reasoning, it is still impossible for the consequent inside that scope to be satisfied. One might take this result to show that, where "ψ" stands for step (3), the antecedents of E3 cannot all be satisfied either – thus using E3 to draw a *modus tollens* rather than rejecting E3 altogether. But I have already argued that the antecedents in E1–E3 can indeed be satisfied: the conjunction of premises (1) and (2) of our paradoxical reasoning and the claim that it implies (3) are both true and knowable; yet conclusion (3) itself is unknowable. There goes principle E3, surely the weakest and (hitherto) most plausible of the standard principles of epistemic closure.

What, one might ask, is so new about that? Haven't principles even as weak as E3 come under fire since J. L. Austin attacked skepticism from an ordinary-language perspective in 1961?[11] Didn't Dretske continue the attack in 1968? Didn't Nozick's later elaboration of Dretske's criticisms put the last nail in the coffin of epistemic closure? (Nozick 1981, Chap. 3). Although the attacks have been fierce, epistemic closure principles have survived, at least according to the champions of closure who continue to publish defenses of it. Peter Klein, for instance, argues that even a principle as strong as E1 emerges completely unscathed from the Dretske–Nozick line of criticism (Klein 1995, 221–3). According to Klein, Dretske and Nozick have confused E1 with a deceptively similar but distinct principle which Klein calls "the mistaken target," and in directing all their fire at the latter principle they have missed E1 completely. So E1, says Klein, remains untouched and may well be true.[12]

But the important point is this: the objections to closure so far raised by Austin, Dretske, Nozick, and others – even if they manage not to aim at a mistaken target – all depend on controversial theories of knowledge. They depend, for instance, on Austin's claim that ordinary language-use is the touchstone and final arbiter in epistemology; or on Dretske's claim that only *some* alternatives to a proposition P are relevant to whether one knows that P; or on Nozick's "tracking" theory, which notoriously fails to contain the KJ Thesis, the claim that knowledge requires justified belief; or, contrariwise, on counterexamples which explicitly *rely* on the KJ Thesis.[13] Thus, a defender of closure can evade those counterexamples by rejecting the epistemological theories or principles on which they depend. My counterexample, however, depends on none of those contentious theories or principles. It takes no side in the debate between ordinary-language epistemologists and Cartesians, between internalists and externalists, between relevant-alternatives epistemologists and their opponents, and so on.

It requires only four plausible assumptions, two which concern the truth-conditions of sentences and two which require merely that what is known to be true be true. Moreover, as I'll argue shortly, it may turn out that we don't even need all four of those assumptions.

4. OBJECTIONS AND REPLIES

4.1. *Relevance Logic*

Someone attracted to relevance logic might object to the presence of material implication in our paradoxical reasoning. But even coming from a relevance logician, that objection would, I think, be mistaken, for only one sentence-constant, "G", appears in the steps of our reasoning, and it appears in each of them – making it hard to see how the premises could fail to be "relevant" to the conclusion. If our relevance logician objects to *any* use of material implication at all, I think our reasoning will go through if we use any other conditional instead, provided that the conditional preserves truth. More to the point, though, this relevance objection is useless to the defender of epistemic closure, who holds that knowledge gets transmitted through known material implication and known strict implication, despite the fact that material implication and strict implication are pariahs in relevance logic.

4.2. *Propositionalism and Sententialism*

A hard-core propositionalist might object that we shouldn't predicate being-known-to-be-true of sentences, such as G, because only propositions, not sentences, can be known to be true. A Quinean "sententialist," on the other hand, might say that only sentences, not propositions, can be known to be true, there being no such things as the propositions envisioned by the propositionalist. Of course, not both of these objections can be right; in any case, I think we can effectively reply to each of them. The propositionalist should accept assumptions A1 and A2, since they are basic to her credo, and she should recognize that assumption A3 predicates being-known-to-be-true, in the first instance, only of propositions and then only derivatively of sentences. Moreover, she can regard the premises of the paradoxical reasoning as themselves propositions, so she should have no objection to their being known to be true. We could satisfy the Quinean objector, on the other hand, this way: (i) Replace assumptions A1–A4 and corollary C with the plausible assumption "If a sentence S is neither true nor false, then any sentence asserting (merely) that S is not known to be true is a true sentence"; (ii) then construe the steps of our paradoxical reasoning,

and the embedded expressions in parentheses, as sentences. Neither our assumptions nor our reasoning would contain talk of propositions, but the reasoning would still go through.

4.3. *Illicit Assumption*

According to the illicit-assumption objection, our paradoxical reasoning relies on an assumption we have no right to make, namely, that we know what G asserts well enough to know what follows from the truth or falsity of G; but *that* assumption, in turn, requires that we know that G expresses a proposition and, furthermore, *which* proposition G expresses. This initially plausible objection is in the end mistaken. Our reasoning does start with the assumption that G expresses a (false) proposition, but that assumption gets discharged in the course of the derivation: the conclusion that G is true does not depend on any undischarged assumption that G expresses a proposition.[14] We do indeed claim to know what would follow if G *were* false, what would follow if G *did* express a false proposition, but the conclusion we derive does not, in the end, depend on the assumption that G *is* false, that G *does* express a false proposition.

4.4. *Truth-Value Gap*

A related objection alleges that sentences like G fall into a truth-value gap, that they are neither true nor false (because, perhaps, they fail to express propositions). If this objection were correct, it would show our reasoning to be unsound, since premise (2) (and its equivalent, (2*)) would come out false. But, like the strengthened version of the Liar Paradox, the Strengthened Knower is specially engineered to block this objection. Suppose the objector is correct: suppose that sentence G is neither true nor false; by corollary C, a consequence of this supposition is that sentence G is *not* known to be true. If the objector is right, what sentence would express this true consequence of her view, namely, that sentence G is not known to be true? Well, among others, sentence G would. So the objector must admit that sentence G is true after all, on pain of having to admit that a sentence expressing a consequence of her view is not a true sentence and, thus, that a consequence of her view is untrue, making the view itself untrue. Since it implies its own negation, the truth-value-gap objection fails.

That rejoinder may strike some people as sophistry, but it seems to me to be a pretty straightforward *reductio* of the assumption that G is neither true nor false. If we really do maintain that sentence G is neither true nor false, then surely we can conclude that G is one sentence that is not known to be true. How might we express the literal truth of this conclusion? Why not via sentence G? But if G expresses a literal truth, it need do nothing

more to be a true sentence. So the assumption that G falls into a truth-value gap leads to a contradiction.

At least one recent author, however, finds such "strengthened" replies to the truth-value-gap objection unpersuasive. Laurence Goldstein maintains that the Strengthened Liar sentence S, "This sentence is not true," does indeed fall into a truth-value gap – and so is not true – but, he says, that result generates no contradiction: "We do say *about* S that it is not true, but that's not 'just what S says it is', since S does not evaluate *itself* as not true; it *fails* to self-evaluate" (Goldstein 1992, 5). Goldstein does not discuss the Knower Paradox, but one might maintain, similarly, both that sentence G falls into a truth-value gap – and so is not known to be true – and also that sentence G does *not* make that very nescience-claim about itself.

Further examination of Goldstein's reasoning, however, suggests that his way out of the Strengthened Liar doesn't fit the Strengthened Knower. According to Goldstein, "the paradox-provoking agent for paradoxes in the Liar family is not negation nor [*sic*] truth nor falsity nor self-reference in general but, as Russell believed, a vicious circle of dependency."[15] Unlike the Strengthened Liar sentence, he says, the self-referential sentence "This sentence is in English" generates no paradox, for the latter sentence "*can* be truth-evaluated, since such an evaluation does not depend on any truth-evaluation, but only on an evaluation, obtained by inspection, of what language the [sentence] is couched in" (Goldstein 1992, 4).

But we can make an analogous claim about the Strengthened Knower sentence, G. We can say, "G *can* be 'knowledge-evaluated,' since such an evaluation does not depend on any knowledge-evaluation, but only on an evaluation, obtained by deduction, of the truth-value of G." Our evaluation of G reveals two things: (i) assuming that G is either true or false, G cannot be false and so must be true, in which case G asserts what is true, in which case G is not known to be true; (ii) assuming that G is neither true nor false, G is not true, in which case, again, G is not known to be true. In each case, we first arrive at a conclusion about the truth-value of G (either G is true, or else G is not true); only then do we arrive at a "knowledge-evaluation" of G, using in the first case the express content of G and in the second case the fact that knowledge implies truth. If I have understood Goldstein's reasoning correctly, sentence G thus avoids the Russellian "vicious circle of dependency" alleged to afflict the Strengthened Liar.

Although, as I've said, the assumption that G is neither true nor false generates a contradiction, the opposite assumption – the assumption that G is either true or false – generates, by contrast, no contradiction at all. In this respect, the Strengthened Knower is quite unlike the Strengthened Liar. In the latter case, one can derive a contradiction both from the assumption that

the Liar sentence is true or false and from the assumption that it is neither true nor false, and, of course, the derivation of a contradiction is the best possible reason for reconsidering one's premises or one's reasoning. But no contradiction follows from assuming that the Knower sentence has a classical truth-value. All that follows is that a certain demonstrably true sentence is unknowable – an unwelcome result, to be sure, but not a self-contradictory one.[16]

The Knower Paradox, then, is in a certain respect special: not every paradox is a counterexample to epistemic closure. The argument of the Sorites Paradox, for instance, leads us to conclude that there are no heaps of sand or, contrariwise, that any collection of grains of sand can make a heap of sand. Since we accept neither of those conclusions, we must regard the argument as invalid or reject at least one of its premises. The same lesson applies *a fortiori* to the Liar Paradox, Russell's Paradox, Grelling's Paradox, the Barber Paradox, and the like: the arguments of those paradoxes generate self-contradictory conclusions and so cannot possibly be sound. Not so the Knower Paradox; there is no reason, for instance, that proposition (3), the conclusion of our reasoning, cannot be true. Indeed, there may be lots of true but unknowable sentences or propositions: assuming that knowledge of a non-axiomatic arithmetical truth requires the provability of that truth within arithmetic, Gödel showed that there are infinitely many unknowable arithmetical truths. The only thing that would make one regard the conclusion of our reasoning as self-contradictory is one's commitment to some principle of epistemic closure: that sounds, then, like a *reductio* of one's commitment to epistemic closure.

4.5. *Hierarchy Theory*

The last objection to the Strengthened Knower I will consider here resembles the objection most often leveled against the reasoning in the Strengthened Liar, namely, that the truth-predicate and the knowledge-predicate are stratified, that there is a hierarchy of truth-predicates, "$true_0$", "$true_1$", etc., and knowledge-predicates, "$known_0$", "$known_1$", etc. According to the hierarchy theory, we may predicate truth or knowledge of a sentence containing a truth- or knowledge-predicate only if the predicate *we* use has a higher subscript than that of any truth- or knowledge-predicate contained in the sentence; otherwise, our prediction is ungrammatical or meaningless.

Even their defenders admit that "Appeals to hierarchies of predicates not present in [natural language] have an especially *ad hoc* flavor" (Anderson 1983, 348). In the words of one opponent, Patrick Grim, "appeal to hierarchy inevitably appears *ad hoc.* Familiar notions of neither truth

nor knowledge seem to come with anything like subscripts attached, and thus at best hierarchical replies have the air of clever technical impositions rather than fully satisfying philosophical solutions" (Grim 1988, 25). Grim argues, moreover, that the hierarchy approach renders unintelligible – indeed, ungrammatical – the notion of a collection of all the truths there are and the notion of omniscience, for, on the hierarchy approach, such global statements as "All true propositions are true" and "God knows all truths" become, not false, but (grammatically) impossible to assert.[17] These results call into serious question the utility and the philosophical integrity of the hierarchy approach. Any theory that regards the tautology "All true propositions are true" as anything other than boringly true has at least two strikes against it, and any theory that solves the problem of divine omniscience by making omniscience an ungrammatical notion seems to me too quick to be plausible.

But even apart from these objections to the hierarchy approach, I'm not sure that the defender of epistemic closure gains much from such an approach. The hierarchy approach would, presumably, classify as ungrammatical any claim to know the first premise of our paradoxical reasoning, i.e., any claim like

(D) $Ks[$(G is known to be true) $\supset \sim$(G is known to be true)$]$

where all three knowledge-predicates are univocal, where they all have the same (implicit) subscript. My counterexample to epistemic closure does rely on D. If I changed the subscripts so as to make "grammatical" both D and the claim that s knows premise (2), then all I could derive via epistemic closure would be something like

(F) $K_1 s \sim$ (G is known$_0$ to be true),

and F presents no contradiction and so no counterexample to closure. But the standard principles of epistemic closure don't come with subscripts either, and so the least we can say is that such principles do not apply with full generality, i.e., to all propositions; they don't apply, for instance, to propositions containing knowledge-predicates.

I can think of one general principle of epistemic closure that avoids falsification by the Knower Paradox, but I have never seen it proposed or defended, in print or anywhere else:

(E4) $\Box \, \forall s \, \forall \phi \, \forall \psi \, ((Ks\phi \; \& \; Ks(\phi \supset \psi) \; \& \; \Diamond Ks\psi) \supset Ks\psi).$

Call E4 the principle of "closure under known implication of a knowable consequent." It avoids the Knower Paradox, all right, since G is unknowable, but it seems to me a useless principle in at least three ways. First,

it does the skeptic no good at all, since in her argument ψ takes as its value the negation of a skeptical hypothesis, a hypothesis which, she says, *cannot* be known to be false; so she can't use E4 to do any argumentative work. But it does the anti-skeptic no dialectical good either, since assuming antecedently that skeptical hypotheses are knowably false seems just to beg the question against skepticism. Third, and more generally, we will not always be able to establish the knowability of ψ apart from actually knowing that ψ is true; we won't, then, be legitimately confident in using E4 until we have established the very conclusion that E4 is there to help us establish. So E4 seems to me useless even if true; it may be no wonder, then, that I have never seen it anywhere.

5. CONCLUSION

So what if I'm right? So what if even the weakest standard principles of epistemic closure fail? I will conclude by mentioning five consequences of the Knower Paradox and of the failure of closure.

First, and most obviously, no correct epistemic logic can contain an epistemic closure principle even as strong as E3, the weakest such principle anybody has bothered to put forward. Second, assuming that anything demonstrably true is true, and assuming that an omniscient being must know all truths, the Knower Paradox may show that there can be no omniscient being.[18]

Third, it may turn out that we have no way of knowing the falsity of any cleverly designed skeptical hypothesis. How, for instance, could I come to know that I am not a brain in a vat? Presumably, I could not come to know it directly, non-inferentially. I take it that not even as staunch an anti-skeptic as G. E. Moore ever claims that I can know that I am not a brain in a vat without inferring it from something *else* I know.[19] The skeptic has deliberately contrived the brain-in-a-vat scenario so as to make it impossible for me to apprehend its falsity directly (say, by merely opening my eyes and seeing that it is false). But if the only way for me to know I'm not a brain in a vat is for me knowingly to infer it from something else I know, then I cannot know I'm not a brain in a vat unless principle E3, or some principle at least as strong, is true.[20] Thus, if E3 is false, I cannot know I'm not a brain in a vat.

Fourth, and fortunately for me, the skeptic cannot use this item of my ignorance to infer that I know nothing at all about my surroundings. If even E3 is false, then the standard arguments for Cartesian skepticism, which attempt to infer wholesale ignorance from our ignorance of the falsity of skeptical scenarios, are unsound. As Klein persuasively argues, the

standard arguments for Cartesian skepticism depend on principles at least as strong as E3; indeed, he says, the best such arguments – those that do not beg the question in favor of skepticism – depend on what he calls the "Immunity Principle," a principle stronger than any of the principles I have attacked here.[21] Thus, if E1–E3 are false, so too is the Immunity Principle, making even the best skeptical arguments unsound.[22]

But maybe I've made too much of the Knower Paradox. Maybe the counterexample to closure provided by the Knower is too contrived to have all, or perhaps any, of the dramatic consequences I've been saying it has. For anything I've shown here, why can't we remain confident about epistemic closure in those many cases that don't involve paradoxical sentences like G? Why can't we simply restrict our closure principles in the appropriate ways?[23] The trick, of course, is to specify in an informative way just what features of G make it the sort of thing our closure principles can properly quarantine; it is unfortunately easy to give a specification that is too weak, too strong, or just trivial.

We don't, for instance, want to exclude all and only self-referential sentences from the scope of our closure principles. As with the Liar Paradox, one can concoct a non-self-referential version of the Knower Paradox that threatens epistemic closure, and there are plainly self-referential sentences that pose no threat at all to closure. Excluding *only* self-referential sentences is too weak, since it doesn't avoid the counterexample to closure arising from these two non-self-referential sentences:

(H)   It is not the case that sentence J is known to be true

and

(J)   Sentence H is true.

If I assume that H is either true or false (an assumption I would defend just as I defended the same assumption about G), I can reason as follows. J is false only if J is true, and so I know the truth of J via *consequentia mirabilis*. I know the truth of the conditional $(J \supset H)$ because its negation, $(J \ \& \sim H)$, is a contradiction. Yet I may well not know the truth of H itself, even though I know its truth to follow from J and $(J \supset H)$. In order for me to know that H is true, I require that "Sentence H is true" *not* be known to be true, and there is no guarantee that this requirement will be met: given that I know which sentence goes by the label "H," if I do indeed know that H is true it may well be that "Sentence H is true" *is* known to be true – by me, for instance. So, even though epistemic closure requires it, there is no guarantee that I know that H is true: hence the counterexample.

By the same token, excluding *all* self-referential sentences is too strong, since it prevents anyone from, for instance, coming to know the truth of

(K)   Some self-referential sentences are true

just by knowingly inferring it from her knowledge of the truth of both the self-referential sentence

(L)   Sentence L is in English

and the conditional $(L \supset K)$. But someone might indeed come to know the truth of K in just that way; we don't want our epistemic closure principles to rule out her doing so. Nor, finally, will it help to exclude all and only "paradoxical" sentences, especially if "paradoxical" is read as "providing a counterexample to closure": for that suggestion reduces to the truism that closure holds except in those cases where it doesn't hold – hardly the guidance we were seeking. If I have made too much, then, of the connection between the Knower Paradox and epistemic closure, it is hard to see how one could avoid making as much of it as I have.

Fifth, and finally, a consequence of the failure of epistemic closure whose significance is harder to gauge. Doesn't my argument prove way too much? What about the conclusions I argue for in this paper? Do I know any of them? On what basis do I know them other than by knowing that they follow from premises I know? In arguing for the fallibility of epistemic closure, am I not sawing off the branch on which I and all other philosophers are perched?

I can think of two replies. First, we might take Nozick's attitude toward the failure of closure: it's not as if knowledge *never* gets transmitted via known deduction from known premises; it's just that sometimes it doesn't (Nozick 1981, 230–4). But Nozick's "tracking" theory is highly contentious, and any Nozickian account of the transmission of knowledge via deduction inherits that problem.[24] On the other hand, without a theory, such as Nozick's, telling us when knowledge gets transmitted via deduction and when it doesn't, we risk relying on a fallible epistemic device without understanding the full range of conditions under which it fails. We saw earlier that, however culpable they may seem to be, neither self-reference nor a Russellian "vicious circle of dependency" is what causes closure to fail. But then what does cause it?

Alternatively, we might challenge the importance of knowledge in all of this. Maybe it's enough that the premises of my argument *are* true, that they *do* imply its conclusion, that we justifiedly believe those premises, that we justifiedly believe all of the inferences we draw, and that we justifiedly

believe the conclusion. Indeed, maybe we justifiedly believe the conclusion *because* we justifiedly believe it to follow from premises we justifiedly believe.[25] Why isn't that enough to continue doing philosophy? Well, we may have to content ourselves with this more modest situation. Still, since I would like to be able to rely infallibly on known deduction from known premises, and since I would like to *know* that I am not a brain in a vat, I'd be happier if instead we could solve the Knower Paradox.

NOTES

[1] See Anderson (1983), (1992), and references cited there. Anderson credits the paradox to Kaplan and Montague (1960). See also Grim (1988).

[2] See, e.g., Feldman (1995), Luper-Foy (1987), and Roth and Ross (1990).

[3] As I use the term "paradox," not all paradoxes generate logically self-contradictory conclusions, although of course many of them do. It suffices that a paradox generate an *objectionable* conclusion from seemingly unobjectionable premises via seemingly unobjectionable reasoning. In this sense, both the Descending Sorites and the Ascending Sorites are paradoxes, since they conclude, respectively, that any collection of grains of sand can make a heap of sand or that no collection can – although, of course, neither argument draws *both* conclusions.

[4] See note 2.

[5] The Strengthened Knower uses a sentence analogous to that used in the Strengthened Liar. If the original Knower sentence is the sentence, G*, "Sentence G* is known to be false," the analogous Strengthened Knower Paradox arises from the logically weaker sentence, G**, "Sentence G** is known not to be true." My version of the Strengthened Knower sentence, G, "It is not the case that sentence G is known to be true," is, however, logically even weaker.

[6] One might object that I implicitly relied on epistemic closure when I argued for the knowability of ((1) & (2)) and for the knowability of the implication from ((1) & (2)) to (3). Such reliance, however, would be perfectly legitimate in a *reductio* of epistemic closure: I can, if I need to, assume that closure holds in order to demonstrate that it doesn't hold.

[7] I reproduce closure principles E1–E3 in essentially the form in which they appear in Hales (1995). See his text for citations to critics and defenders of each of these principles.

[8] See, e.g., Hales (1995, 188–192).

[9] Hales (1995, 189, 192) considers this kind of reply and finds it unpersuasive.

[10] See Hales (1995, 192) and references cited there.

[11] Austin (1961, 45), cited in Stroud (1984, 44).

[12] Klein (1995), 214, 221–225. The mistaken target principle (p. 221) is this, where e, x, and y are propositions:

> $(\forall x)(\forall y)$ (If e is an adequate source of S's justification for x, and x entails y, then e [rather than x itself] is an adequate source of S's justification for y).

Klein agrees that the Dretske–Nozick counterexamples threaten the mistaken target principle, but, he argues, they leave standard epistemic closure principles, such as E1, untouched. I agree with Klein's criticism of the Dretske–Nozick attack but not, of course, with his claim that E1 may well be true.

[13] For an account that agrees with Nozick (and Dretske) at least in rejecting the KJ Thesis, see Alston (1989). For a counterexample to E2 which explicitly relies on the KJ Thesis, see Hales (1995), 192–193.

[14] In a natural-deduction representation of our reasoning, the conclusion that G is true would lie outside the scope-line of the assumption that G is false. Compare Anderson (1992, 324).

[15] Goldstein (1992, 4). Compare Haack (1978, 141–151).

[16] It would be a self-contradictory result only if we defined "demonstration" as a process which infallibly transmits knowledge from known premises via know-ledge-preserving inferences. But that move would beg the question (or else commit the fallacy of persuasive definition) here, since it would make demonstration impossible in the absence of epistemic closure.

[17] Compare Grim (1988, 25).

[18] Compare Grim (1988, 15) for a different use of the Knower against omniscience.

[19] This inferential strategy for knowing the falsity of skeptical scenarios seems to be the one Moore favors in, e.g., Moore (1959).

[20] At any rate, the truth of a principle at least as strong as E3 – i.e., a universal generalization covering all epistemic subjects and all objects of knowledge – seems to be the only *non-arbitrary* explanation of my coming to know that I'm not a brain in a vat.

[21] Klein (1995), 229–230 nn. 3, 6, and references cited there. By "wholesale ignorance," I mean ignorance of every proposition entailing the existence of an external world. Some philosophers argue that "semi-skepticism" – narrower in scope than Cartesian skepticism yet wider in scope than the skeptical concessions of Nozick and his fellow travelers – can be defended without recourse to epistemic closure; see Jonathan Vogel, "Are There Counterexamples to the Closure Principle?" in Roth and Ross (1990, 13–27, especially 21–22).

[22] This outcome – namely, that I cannot know the falsity of skeptical scenarios even though the standard arguments for Cartesian skepticism are unsound – supports the Dretske–Nozick line on skepticism, and, indeed, another consequence of the Knower Paradox is that the Dretske-Nozick line on skepticism may be right after all.

[23] For pressing me about this possibility, I'm indebted to a number of people, in particular Bruce Hunter, Tom Vinci, and an anonymous referee for *Synthese*.

[24] What is worse, Nozick's theory does not explain why closure fails in the case of the Knower Paradox. According to Nozick, s knows that Q via known deduction from P if and only if (i) P is true; (ii) Q is true and inferred by s from P; and these two counterfactual conditionals also hold: (iii) $\sim$Q $\square\rightarrow$ [$\sim$(s believes that P) $\vee$ $\sim$(s infers Q from P)]; (iv) Q $\square\rightarrow$ (s believes that P). Let P be the conjunction of steps (1) and (2) of our paradoxical reasoning, and let Q be G. In that case, conditions (i) and (ii) are easily met; (iii) holds

trivially in virtue of an impossible antecedent (since G cannot possibly be false), and so (iii) "drops out" of the analysis – leaving condition (iv), which does seem to be satisfied: in the nearest worlds in which Q is (still) true, *s* (still) believes P on the basis of the reasoning used to establish ((1) & (2)). Since Nozick's conditions for the transmission of knowledge are satisfied in this case, his theory does not explain the "transmission failure" brought about by the Knower Paradox. (See Nozick 1981, 186, 231–234.)

[25] I can't envision a version of the Knower Paradox that threatens the analogous doxastic closure principle – the closure of justified belief under justifiedly believed entailment – simply because, unlike knowledge, justified belief (at least as ordinarily understood) doesn't imply truth. If the Lottery Paradox strikes you as a counterexample to the doxastic closure principle, compare Klein (1995): "[The reasoning that generates the Lottery Paradox] does not provide *any* evidence against the Closure Principle, since that principle concerns the transmission of justification from *one* proposition to another through entailment; it does not concern the transmission of justification from *sets* of justified propositions containing more than one member to another proposition entailed by that set" (218, emphasis in original).

REFERENCES

Alston, W. P.: 1989, 'Justification and Knowledge', *Epistemic Justification: Essays in the Theory of Knowledge*, Cornell University Press, Ithaca and London, pp. 172–182.

Anderson, C. A.: 1983, 'The Paradox of the Knower', *Journal of Philosophy* **80**, 338–355.

Anderson, C. A.: 1992, 'Paradox of the Knower', in Jonathan Dancy and Ernest Sosa (eds.), *A Companion to Epistemology*, Blackwell, Oxford, pp. 324–326.

Austin, J. L.: 1961, 'Other Minds', *Philosophical Papers*, Oxford University Press, Oxford.

Dretske, F.: 1968, 'Reasons and Consequences', *Analysis* **28**, 166–168.

Dretske, F.: 1970, 'Epistemic Operators', *Journal of Philosophy* **67**, 1007–1023.

Dretske, F.: 1971, 'Conclusive Reasons', *Australasian Journal of Philosophy* **49**, 1–22.

Feldman, R.: 1995, 'In Defence of Closure', *Philosophical Quarterly* **45**, 487–499.

Goldstein, L.: 1992, ' "This Statement is Not True" is Not True', *Analysis* **52**, 1–5.

Grim, P.: 1988, 'Truth, Omniscience, and the Knower', *Philosophical Studies* **54**, 9–41.

Haack, S.: 1978, *Philosophy of Logics*, Cambridge University Press, Cambridge.

Hales, S. D.: 1995, 'Epistemic Closure Principles', *Southern Journal of Philosophy* **33**, 185–201.

Hintikka, J.: 1962, *Knowledge and Belief*, Cornell University Press, Ithaca, N.Y.

Kaplan, D. and R. Montague: 1960, 'A Paradox Regained', *Notre Dame Journal of Formal Logic* **1**, 79–90.

Klein, P.: 1995, 'Skepticism and Closure: Why the Evil Genius Argument Fails', *Philosophical Topics* **23**, 213–236.

Luper-Foy, S. (ed.): 1987, *The Possibility of Knowledge: Nozick and His Critics*, Rowman and Littlefield, Totowa, N.J.

Moore, G. E.: 1959, 'Certainty', *Philosophical Papers*, George Allen and Unwin, London, pp. 227–251.

Nozick, R.: 1981, *Philosophical Explanations*, Harvard University Press, Cambridge, Mass.

Roth, M. and G. Ross (eds.): 1990, *Doubting: Contemporary Perspectives on Skepticism*, Kluwer Academic Publishers, Dordrecht.

354 STEPHEN MAITZEN

Stroud, B.: 1984, *The Significance of Philosophical Scepticism*, Oxford University Press, Oxford.

Department of Philosophy
Dalhousie University
Halifax, Nova Scotia B3H 3J5