

RESTORATIVE RIGGING AND THE SAFE
INDICATION ACCOUNT

ABSTRACT. Typical Gettieresque scenarios involve a subject, S, using a method, M, of believing something, p, where, normally, M is a reliable indicator of the truth of p, yet, in S's circumstances, M is not reliable: M is deleteriously rigged. A different sort of scenario involves rigging that restores the reliability of a method M that is deleteriously rigged: M is restoratively rigged. Some theorists criticize (among others) the safe indication account of knowledge defended by Luper, Sosa, and Williamson on the grounds that it treats such cases as knowledge. But other theorists also criticize the safe indication account because it treats the cases as examples of ignorance when they are really examples of knowledge. I answer both groups of critics by arguing that (1) restorative rigging can enable us to know things, and (2) restoratively rigged cases can meet the relevant conditions of the safe indication account.

Typical Gettieresque scenarios involve a subject, say S, who uses a method, M, of acquiring (or sustaining) a belief, p, where, normally, M is a reliable indicator of the truth of p, yet, in S's peculiar circumstances, M is not reliable. Let us say that in such circumstances M has been *deleteriously rigged*. Recently, epistemologists have discussed a different sort of scenario, involving rigging that restores the reliability of a method M that is deleteriously rigged. We might say that in such scenarios M has been *restoratively rigged*. Some theorists, including Robert Shope (2002; also Joe Salerno, unpublished) have adduced cases of restorative rigging against (among others) the safe indication account of knowledge defended by Steven Luper (1984, 1987, 2002), Ernest Sosa (2002, 2003), and Timothy Williamson (2000). Restoratively rigged cases have at least some of the earmarks of Gettier's original examples. Shope thinks that restorative rigging is just another layer of Gettieresque manipulation, so that examples involving such manipulation are Gettier cases. He then criticizes the safe indication account on the grounds that it treats these examples as if they involved genuine knowledge. But (at least) some restoratively rigged cases strike other epistemologists, such as Neta and Rohrbaugh (2004) and Juan Comesaña

(2004), as examples of genuine knowledge, and hence not as Gettier scenarios at all. These theorists also criticize the safe indication account – this time because cases of restorative rigging are really examples of knowledge but cannot meet the central condition of the safe indication account, which is roughly that, at time *t*, *S* knows *p* by arriving at the belief *p* through some method *M* only if:

(T:) *M* would, at *t*, indicate that *p* was true only if *p* were true.

If I have understood them, it cannot be that both groups of critics are correct (about whether restoratively rigged cases are Gettieresque). But both can be mistaken (about whether these cases are counterexamples to the safe indication account). I will argue, as against Shope and like-minded critics, that (1) restorative rigging is not a form of Gettierizing; restorative rigging can enable us to know things. Against Neta and Rohrbaugh, and similar critics, I will argue that (2) restorative rigging does not bar a case from meeting the safe indication condition T. So neither group of critics has provided grounds for rejecting the safe indication account after all. I begin with (1).

The initial plausibility of (1) will seem clear to epistemologists who adopt the widely accepted externalist view that my knowing *p* via some method *M* requires that *M* be reliable in my circumstances. For, as we have observed, a method's reliability can be impaired by one set of factors yet restored by another. If *M*'s reliability can be ensured by restorative rigging, and knowing *p* via *M* hinges on *M*'s reliability in my circumstances, then restorative rigging can be the critical factor in whether I know *p* via *M*. It can enable us to meet a condition that is necessary for knowledge.

By appealing to examples, we can bring out the acceptability of the implications of (1). Let us assume that the following is a straightforward case of knowledge:

- (1) *Ordinary Barn Case*. *S* stands before a barn and forms a visual impression of it in optimal viewing conditions, coming thereby to believe a barn is there. *S*'s belief is not based on any explicit reasoning. *S* does not, for example, appeal to other beliefs in coming to *S*'s view. In fact, *S* has little or no idea about how vision works.

Let us also assume that the following familiar case (from Alvin Goldman 1976) is an example of ignorance (it involves deleterious rigging):

- (2) *Barn Case with Deleterious Rigging.* S stands before the only real barn in a neighborhood that, unbeknownst to S, is filled with convincing papier-mâché look-alikes. S has not seen the copies; instead, as in the Ordinary Barn Case, S forms a visual impression of the real barn and in that way, using no explicit reasoning, believes that a barn is there.

To this case we might add one detail: S never will see one of the copies, due perhaps to S's preoccupation with pressing tasks and the route S takes out of the neighborhood. Hence S's seeing a facsimile is, as things turn out, a wholly counterfactual possibility. Clearly this detail changes nothing as to S's possessing knowledge: no one would say that, while looking at the real barn, S knows it is there so long as S never actually sees any of the fakes but otherwise S does not. One can imagine many additional situations in which the reliability of a method is undermined by close yet counterfactual possibilities, precluding the transmission of knowledge.

But now consider a third case.

- (3) *Case with Restorative Rigging.* This is just like 2 except that, during this time of year, X-smog (as we might call it) always fills the part of the world S occupies. X-smog has a peculiarity: in a completely reliable way, it conceals all papier-mâché barns but no real barns, thus stopping any fake barns from causing in S barn-type visual percepts which the fakes might otherwise have produced.

I suggest that, in this third case, as in the first, S knows a barn is there. Given the circumstances portrayed in the *second* – deleteriously rigged – case, visual impressions of barns are unreliable indicators that barns are present, since such impressions might be produced by fake barns. This is why the second case clearly involves knowledge failure. But in our latest example the reliability of barn percepts (as indicators of the presence of barns) has been restored – it is false, given the circumstances, that barn percepts might have been caused by fake barns, and true that they would only be caused by real barns. Hence there is no failure of knowledge in our third case. (The suggestion that there is no failure of knowledge in the third example can be reinforced if we notice that, in a great many straightforward cases of knowledge – cases that do not seem Gettieresque – there is restorative ‘rigging.’ That is, in many quite straightforward cases, while there are factors that tend to undermine

the reliability of belief formation methods, other factors counteract the first.)

Using the Stalnaker–Lewis (1968, 1973) semantics for subjunctive conditionals (supplemented with Robert Nozick’s gloss in 1981, 174, no 8) we can put the point this way: the nearest worlds in which S is deceived by barn simulacra are significantly farther from the actual world as it is portrayed in scenario 3 than they are in the actual world as it is portrayed in scenario 2. Figuratively speaking, the deleterious rigging in 2 ‘pulls’ worlds in which S is deceived closer in than they are in scenario 1, while the restorative rigging ‘pushes’ them back out, although the restorative rigging does not ‘push’ them as far out as they are in 1. Still, they are far enough: there is knowledge failure in neither 1 nor 3.

In case 3, the restorative rigging was contemporaneous with the deleterious rigging. But knowledge is possible even in cases in which the former occurs after the latter. For example, the description of our first case could easily be expanded as follows without changing our intuition that in it S knows a barn is there:

- (4) *Case with Noncontemporaneous Restorative Rigging.* An artist is about to implement a plan to spread papier-mâché barns throughout the neighborhood in which S will later walk. The fakes are loaded in a truck, ready to go. But the artist has several friends who think such projects are absurd, and convince him to change his mind. So no fakes are in the neighborhood. It is still later that S enters the area, stands before a barn, and forms a visual impression of it in optimal viewing conditions, coming thereby to believe a barn is there.

The artist’s preparations constitute a factor that, while in place, tends to undermine the reliability of S’s method, by raising the likelihood of S’s exposure to barn duplicates. But the friends’ determination to convince the artist not to make a fool of himself, and their success, constitute a factor that, by the time S believes there is a barn, restores the reliability of S’s method, by once again eliminating the prospect of S’s exposure to replicas. Because of the friends’ actions, it is false that, at the relevant time, S’s barn percepts might have been caused by a fake barn.

We can distinguish (a) manipulation occurring in a scenario we design from (b) manipulation occurring in the design of a scenario. All of our examples so far involve (b), and none (a). That is, there is no character *appearing in* any of our four scenarios

who deleteriously rigged S's belief-formation method. Nor is there anyone who restored the reliability of that method. But nothing important hinges on this distinction. Suppose that the reliability of S's deleteriously rigged method is not restored by an impersonal factor in S's situation, as the fog does in our third case, but rather its reliability is restored by a person or agent deliberately, as in the following example:

- (5) *Case with Deliberate Restorative Rigging.* This is like 2 except that a team of scientists has decided to ensure that S will not fall into error when S relies on visual impressions of barns. The team does nothing while S examines the real barn, but stands ready to prevent S from forming any visual impressions of the fakes. They are carefully monitoring S's eye movements and will completely destroy any replica that is about to come into S's line of vision just before S sees it.

In the scenario itself there is an agent who deliberately fashions the rigging so as to restore the reliability of a method of belief formation. We might say it involves *deliberate* restorative rigging. However, in spite of the deliberate rigging, it is as much a case of knowledge as 3 and 4. Given the circumstances as portrayed in 3 and 4 as well as 5, it is false that barn percepts might have been caused by fake barns, and true that they would only be caused by real barns. The only difference between this latest case and the other two is that in it someone is *trying* to make S's method of belief formation reliable. Why should *that* prevent S from attaining knowledge? People take all sorts of steps, many of which we know nothing about, to ensure that we have reliable sources of information about things, and much of the time they thereby enable us to know things.

(Conceivably, the restorative rigging of a method might be undermined by further deleterious rigging. We can treat such situations as we do cases of simple deleterious rigging, and say they involve knowledge failure. Too, when the restorative rigging of a method is undermined by deleterious rigging, the deleterious rigging might itself be reversed by further restorative rigging (and so forth). We can treat such scenarios as we do cases of simple restorative rigging, and say they do not involve knowledge failure.)

This concludes the positive defense of my claim that restorative rigging can enable us to know things. Eventually I will consider some objections, but first I will respond to Shope's attempt to offer such cases as counterexamples to the safe indication account.

Shope assumes that cases of restorative rigging involve knowledge failure. For example, Shope says the following case is one of knowledge failure (2002, 37):

- (6) *Guardian Angel Case*. This is just like the Deleteriously Rigged Case except that “a guardian angel is present who would block the formation of a false belief in S that [there is a barn], were S to look toward a mere facsimile, for example, by blurring S’s vision or by stopping S’s sensory experience from causing S to believe [there is a barn].”

Shope offers his scenario as a counterexample to Goldman’s (1976) version of reliabilism, but he also suggests that the safe indication account is vulnerable to examples involving “a mere rigging by external manipulators of a match between S’s beliefs and the facts (47),” and I presume that the Guardian Angel Case is just such an example. But Shope’s case is essentially the same as our last example – case 5. It is a scenario that straightforwardly involves deliberate restorative rigging, and hence it is not a counterexample at all.

Now let us consider whether restorative rigging bars a case from meeting the safe indication condition T. It is clear that, in all of our examples except for the second, condition T is met. In our scenarios, S would have the percepts that indicate the presence of a barn only if there were a barn. So restorative rigging does not prevent a method from meeting the key condition of the safe indication account.

Yet Neta and Rohrbaugh offer cases of restorative rigging as examples of knowledge that are ruled by the safe indication account to be ignorance. Here are the main examples they discuss:

- (A) I am drinking a glass of water which I have just poured from the bottle. Standing next to me is a happy person who has just won the lottery. Had this person lost the lottery, she would have maliciously polluted my water with a tasteless, odorless, colorless toxin. But since she won the lottery she does no such thing. . . Now, I drink the pure, unadulterated water and judge, truly and knowingly, that I am drinking pure, unadulterated water.
- (B) I am participating in a psychological experiment, in which I am to report the number of flashes I recall being show. Before being shown the stimuli, I consume a glass of liquid at the request of the experimenter. Unbeknownst to either of us, I have been randomly assigned to the control group, and the glass con-

tains ordinary orange juice. Other experimental groups receive juice mixed with one of a variety of chemicals which hinder the functioning of memory without a detectable phenomenological difference. I am shown seven flashes and judge, truly and knowingly, that I have been shown seven flashes.

I think most of us will be somewhat inclined to treat A and perhaps even B as cases of knowledge. That is because we tend to think A and B involve restorative rigging. Concerning A we see that, while my circumstances are such that I am subject to being poisoned by my enemy, my usual method of detecting water is unreliable, since I cannot distinguish water from the poison. (Compare Case 2, in which S is liable to see a fake barn.) But we reason that, after having won the lottery, my adversary would not poison me; hence, in my new circumstances, it is no longer true that I might have been poisoned. My enemy's new frame of mind restores the reliability of my usual method of detecting non-toxic water, by making it the case that my water percepts would not be caused by toxic water, so that, were I to believe I'm drinking water on the usual basis, I would be correct. Similarly with B. While I am subject to being given the memory altering chemicals, my circumstances are such that in them my memory is not reliable. But we reason that the decision not to give me the tainted water restores the reliability of my memory, making it the case that, in my circumstances as they are after it is decided that I will not be given the tainted water, my judgment about the seven flashes would not be caused through chemical manipulation.

I have conceded that A and B can appear to be cases of knowledge. But Neta and Rohrbaugh say that (1) A and B are *clear* cases of knowledge and that (2) the safe indication account clearly does not count them as cases of knowledge.

As to Neta and Rohrbaugh's first point: what happens in A and B is *borderline* restorative rigging. In our cases 3–5, deleterious rigging (that 'pulls' worlds in which S is deceived closer to the actual world) is neutralized by restorative rigging that 'pushes' the deceptive worlds back out. In cases A and B, factors also (non-contemporaneously) neutralize deleterious rigging to some extent, but these factors do not 'push' the deceptive worlds very far from the actual world. For this reason, whether A and B involve knowledge failure is no straightforward matter, and there is room for reasonable disagreement. Let us add that, even if A and B were

paradigm cases of restorative rigging, it still would be a mistake to classify them as *clear* cases of knowledge, for some theorists say, even of clear cases of restorative rigging, that they are cases of *ignorance*. Apparently, Shope takes this view. The most straightforward explanation of his reaction to case 6 is that he thinks knowledge cannot be produced through restorative rigging – that cases involving deleterious rigging remain cases of ignorance even if restoratively rigged. Hence, it would seem, there are varying intuitions about whether restorative rigging can produce knowledge.

There is another way to argue that in A (and B) knowledge exists: one might say that, although I cannot discriminate between pure and tainted water on the basis of taste (and related cues), I still know the water by its taste, since my drinking the tainted water is a purely counterfactual possibility, and purely counterfactual possibilities cannot stop a method from transmitting knowledge. But we have already established that this claim is false. In our discussion of case 2 we noted that S's epistemic status does not hinge on S's never actually seeing a simulacrum: even though S's seeing a fake is a counterfactual possibility, it is enough to undermine S's knowledge claim. Wholly counterfactual possibilities are quite capable of blocking the transmission of knowledge. (With a little effort we can imagine many further ways to illustrate the point that close yet counterfactual possibilities can preclude the transmission of knowledge. For example, as in 1, S sees a real barn in a neighborhood that is free of fakes, but hovering nearby is a demon who can quickly and easily create convincing fakes or sensory illusions, and who is thinking about whether to do so, having not yet decided. Isn't the close possibility that the demon will act sufficient to undermine the reliability of S's method of belief formation and to call S's claim to knowledge into question?)

As to the second claim made by Neta and Rohrbaugh: it is *not* completely clear that A and B pass the bar for knowledge according to the safe indication condition T. Its proponents have some wiggle room. Employing T, we can explain why people might find any case of restorative rigging hard to classify, especially a borderline case such as A or B. T can be explained as follows: p is true in all worlds close to the actual world in which S's method M indicates that p is true. In this explanation of T, the term 'close' is, of course, imprecise. Certainly T is met in case 1 and not in 2. It is less clear that it is met in cases of restorative rigging; however, I have suggested that the deceptive worlds involved in restorative rigging

are far enough from the actual world: restorative rigging upholds knowledge. Neta and Rohrbaugh trade on examples that are borderline cases of restorative rigging. Are the deceptive worlds in Neta and Rohrbaugh's examples close enough for T to be met? How close is close enough? The matter is not determinate. Hence it is not clear whether they meet T. But this fact counts in favor of the safe indication account, since the examples are themselves not intuitively clear cases.

I will close by responding to an objection, which is this: none of examples 1–6 involves genuine knowledge, for in none of them does S have good reason for thinking that S's method of belief formation is reliable since S is wholly unaware of the factors that determine whether S's method of belief formation is reliable. The presence or absence of rigging, whether deliberate or not, is merely part of the environment in which S operates. For all S knows, factors that tend to undermine the reliability of S's methods are not cancelled by factors that restore the reliability of S's methods.

This is the standard objection which internalist epistemologists make against all of the proposals of externalist epistemologists. Externalists suggest that we can know things by virtue of meeting various conditions we have no idea we meet; internalists (most notably Lawrence BonJour) reply that the factors that make for knowledge must be revealed from within the cognitive perspective of the knower, or else we will be correct only 'by accident,' and our belief will be 'epistemically irresponsible.' Admittedly, I have said nothing here to resolve this debate. Instead, I have simply assumed that internalism is false, and addressed my argument to externalist-minded epistemologists. We all know why internalists will deny that knowledge is sustained by deliberate restorative rigging. What I have argued is that externalists have no good reason to deny this.

ACKNOWLEDGEMENTS

I thank my colleagues, and *Synthese's* anonymous reviewer, for helpful comments on an earlier version of this essay.

REFERENCES

- BonJour, L.: 1980, 'Externalist Theories of Empirical Knowledge', *Midwest Studies in Philosophy* 5, 53–73.

- Comesaña, J.: 2004, 'Unsafe Knowledge', Forthcoming in *Synthese*.
- Dretske, F.: 1971, 'Conclusive Reasons', *Australasian Journal of Philosophy* **49**, 1–22.
- Goldman, A.: 1976, 'Discrimination and Perceptual Knowledge', *Journal of Philosophy* **73**, 771–791.
- Goldman, A.: 1979, 'What Is Justified Belief?', in G. Pappas (ed.), *Justification and Knowledge*, Reidel, Dordrecht, pp. 1–23.
- Lewis, D.: 1973, *Counterfactuals*, Harvard University Press, Cambridge.
- Lewis, D.: 1996, 'Elusive Knowledge', *Australasian Journal of Philosophy* **47**(4), 549–567.
- Luper(-Foy), S.: 1984, 'The Epistemic Predicament: Knowledge, Nozickian Tracking, and Skepticism', *Australasian Journal of Philosophy* **62**, 26–48.
- Luper(-Foy), S.: 1987, 'The Causal Indicator Analysis of Knowledge', *Philosophy and Phenomenological Research* **47**, 563–587.
- Luper(-Foy), S.: 2002, 'Indiscernability Skepticism', in S. Luper (ed.), *The Sceptics*, Ashgate Publishing Limited, Aldershot, pp. 183–203.
- Neta and Rohrbaugh: 2004, 'Luminosity and the Safety of Knowledge', Forthcoming in *Pacific Philosophical Quarterly*.
- Nozick, R.: 1981, *Philosophical Explanations*, Harvard University Press, Cambridge.
- Shope, R.: 2002, 'Conditions and Analyses of Knowledge', in P. Moser (ed.), *The Oxford Handbook of Epistemology*, pp. 25–71.
- Sosa, E.: 2002, 'Tracking, Competence, and Knowledge', in P. Moser (ed.), *The Oxford Handbook of Epistemology*, pp. 264–287.
- Sosa, E.: 2003, 'Neither Contextualism nor Skepticism', in S. Luper (ed.), *The Sceptics*, Ashgate Publishing Limited, Aldershot, pp. 165–183.
- Stalnaker, R.: 1968, 'A Theory of Conditionals', in N. Rescher (ed.), *Studies in Logical Theory*, Basil Blackwell, Oxford.
- Williamson, T.: 2000, *Knowledge and Its Limits*, Oxford University Press, Oxford.

Philosophy Department
Trinity University
715 Stadium Dr.
San Antonio, TX 78212-7200
USA
E-mail: sluper@trinity.edu