Adam Leite
Fallibilism

In the broadest sense of the term, *fallibilism* is an anti-dogmatic intellectual stance or attitude: an openness to the possibility that one has made an error and an accompanying willingness to give a fair hearing to arguments that one's belief is incorrect (no matter what that belief happens to be about). So understood, fallibilism's central insight is that it is possible to remain open to new evidence and arguments while also reasonably treating an issue as settled for the purposes of current inquiry and action. Fallibilism, so construed, was given its most influential formulation – and its name – by C. S. Peirce, though it was advocated by earlier philosophers as well, particularly the later ancient skeptic Philo of Larissa and perhaps also Hume. Contemporary epistemologists almost universally agree in endorsing this intellectual stance; it is part of the undisputed framework within which contemporary epistemological theorizing takes place.

In some recent discussions, the term "fallibilism" stands for the thesis that human beings are fallible about everything (or just about everything) they believe (Haack 1979). More commonly, however, "fallibilism" is used as a name for a thesis about knowledge and justification: that we can have fallible justifications for our beliefs, and that it is possible to know that something is the case even if one has only a fallible justification for believing it. Differences of formulation and emphasis aside, this thesis is widely accepted, though there is also significant dissent. The following article will focus on issues relating to fallibilism understood in this latter sense.

According to standard usage, a *fallible* justification is one which is compatible with the falsity of the belief in question, in that having this justification does not necessitate or entail the truth of the belief. A justification can thus be said to be fallible even if there is a law of nature to the effect that whenever someone has a justification of that type, then if the person had the belief in question, it would be true. Assuming that knowledge requires justified belief, fallibilism about knowledge can be formulated as follows:

For some P, it is possible for one to know that P even if one could have
exactly the same justification for believing P when P is false.

This thesis is sometimes formulated in terms of the notion of *evidence*:

For some P, it is possible for one to know that P even if one's evidence for
P does not necessitate or entail the truth of P.

However, this latter formulation is unnecessarily narrow, since a fallibilist could deny that knowledge requires justification by evidence.

According to another influential formulation, fallibilist views of knowledge allow that one can know that p even if one is not in a position to "eliminate" or "rule out" every possible alternative to the truth of p (Lewis 1996). This characterization is not clearly equivalent to the previous one. If we treat "eliminate" and "rule out" as requiring a position which *entails* or *necessitates* the falsity of the possible alternative in question, then the two characterizations are arguably equivalent. However, if "eliminate" or "rule out" just means "know not to obtain" (as Dretske (1981) proposes), then this characterization of fallibilism is not equivalent to the previous one. Fallibilists (in the previous sense) about knowledge will treat "ruling out" or "eliminating" as a fallible relation. They will

consequently hold that one's position regarding p can be fallible *even if* one is able to rule out or eliminate every possible alternative to the truth of p. For this reason, they can hold that knowledge requires that one be in a position to rule out or eliminate every possible alternative to the truth of one's belief.

Some externalist or reliabilist accounts of knowledge dispense with any justification requirement altogether. However, such accounts can still be fallibilist in a closely related sense of the term. Any account of knowledge will require that the belief in question be true. It will also require the satisfaction of certain other requirements, some having to do with the person and his or her beliefs, some having to do with the surrounding circumstances. We can use the term "epistemic position" as a label for the person's position in relation to these latter requirements. An account of knowledge can then be said to be fallibilist if it allows that one can have knowledge even if one's epistemic position does not entail the truth of the belief in question – that is, even if satisfaction of those further requirements (over and above the truth and the belief requirements) does not necessitate or entail the truth of the belief.

These differences of formulation create some difficulty in determining whether certain theories of knowledge are fallibilist or not. In the end, however, it is relatively unimportant whether one calls a given theory "fallibilist" or "infallibilist." What matters is that one is clear about what exactly one means in calling it one thing or the other.

Fallibilism faces a further important problem of formulation. It is a commonplace in recent discussions of a priori knowledge that we can have *fallible* justification for believing necessary truths (Bonjour 1997, Casullo 2003). However, this commonplace cannot be accommodated within the standard characterization of fallibilism. A necessary truth can't be false and is (on standard accounts) entailed by everything. It consequently is not possible to believe a necessary truth on the basis of a justification which does not entail its truth.

Recent attempts to solve this problem have suggested that one has fallible knowledge that p if one's belief that p could have failed to be knowledge – either because it could have been false or for some other reason (Hetherington 1999, Reed 2002). This response is not fully satisfying. Epistemologists who speak of fallible a priori knowledge have in mind the possibility that one could know something a priori even though there is, in some sense, a gap between one's evidence or justification and the truth of the known proposition. No one has yet fully explained what it means to talk of a gap here.

Motivations for fallibilism
External world skepticism has provided a significant motivation for the widespread acceptance of fallibilism about knowledge of the external world. Since – it is said – we cannot attain an infallible position regarding most or all of what we believe about the world, many epistemologists have held that infallibilism leads to skepticism, and so they have taken fallibilism to be the only acceptable option. Some epistemologists (Nozick 1981, Dretske 1970, 1981) have taken skepticism to arise from the principle that knowledge is closed under known entailment (roughly: if you know p and you know that p implies q, then you

know q), and so have taken denial of this principle to be necessary to a fallibilist response to skepticism. The resulting view holds that we can know things about the world without knowing that we are not dreaming, not brains in vats, etc. However, a fallibilist response to skepticism need not require denial of the closure principle. A fallibilist could accept, for instance, that in order to know anything about the world, one must know that one is not dreaming; the fallibilist will simply have to explain how one can have fallible knowledge that one is not dreaming.

It has recently been argued that an infallibilist conception of knowledge does not generate external world skepticism because we can have infallible justifications for our beliefs about the world (McDowell 1994, 1995, Neta 2002). For instance, you might justify your belief that there is a table in front of you by claiming that you see that there is a table in front of you, which of course entails that there *is* a table in front of you. However, it is doubtful that this move succeeds by itself. From the viewpoint of the infallibilist skeptic, the question will simply become what justification you have for believing that you *see* that there is a table there, and it is doubtful that you can provide an infallible, non-question-begging justification for this belief.

Even if we can have infallible justifications for some beliefs, a second motivation for fallibilism comes from consideration of ordinary cases of knowledge. Most of our scientific knowledge, and much of our everyday knowledge, seems to arise through reasoning and inference from evidence that does not entail or necessitate the truth of our beliefs. Attempts to interpret such cases within an infallibilist framework are bound to seem forced. Similarly, in ordinary circumstances it is indisputable that one knows who is President, but it is hard to see how one could ordinarily be in an infallible position regarding such matters. (Your epistemic position usually doesn't entail or necessitate that the President didn't die in the past five minutes.) In the end, careful attention to the details of our ordinary practices of knowledge attribution provides the best basis for accepting fallibilism (Austin 1961, Leite 2004).

Objections, Problems, and Prospects

It is sometimes thought that on a fallibilist view, one cannot rationally assert or conclude "p", but only something like, "my evidence supports p" or "so far as I know, p." However, this criticism is incorrect. The fallibilist view is that one can know the truth of a proposition on the basis of grounds that do not necessitate its truth. Consequently, even if there is a norm to the effect that one can't properly assert p unless one knows that p, fallibilism does not imply that one cannot rationally assert or conclude "p."

It is likewise sometimes thought that a fallibilist view precludes us from being able to know that we know that p. This is because on a fallibilist view, there will be no introspectible difference between a case in which one has knowledge and some possible case in which one does not have knowledge. However, this worry arises from a certain preconception of what would be involved in knowing that one knows. An appropriately designed fallibilist theory of knowledge will not preclude us from knowing that we know, so long as the theory is consistently applied at both the first and second order (Feldman 1981).

The slogan "If you know p, then you can't be wrong" has been taken to raise particular problems for fallibilism (Austin 1961). However, this slogan is susceptible to multiple interpretations. If we interpret it as, "Necessarily, if one knows p, then p is true," then it simply states that knowledge requires truth, which fallibilism does not deny. If we interpret it as asserting that when one knows that p, there is a tight causal, nomological, or counterfactual connection between one's belief state and the truth, then fallibilism need not deny it. If we interpret it as asserting that whenever one knows that p, there must be something about one's belief state that entails or necessitates the truth of one's belief, then of course fallibilism denies it. But so interpreted, the slogan is hardly a truism.

Some philosophers (Austin 1961, Kaplan 2006) have worried that fallibilism licenses apparently incoherent assertions such as "I know that p, but I might be mistaken." The worry arises as follows. Fallibilism licenses one to assert "I know that p" when one's epistemic position regarding p does not entail or necessitate p's truth. In such circumstances, one's epistemic position is compatible with possible ways in which one could be wrong about p. Recognizing this, it seems that one should be able to assert "I might be mistaken about p." Unfortunately, putting the two together results in an assertion that seemingly both claims knowledge and repudiates that very claim.

Some fallibilists have suggested that the resulting assertion's oddness can be explained away by appealing to the pragmatic (non-semantic) commitments one makes when one claims to know something (Austin 1961, similar ideas in Rysiew 2001). However, a fallibilist doesn't have to grant that such an assertion would ever be true or appropriately made. The term "might" in the problematic assertion is most naturally interpreted in terms of epistemic possibility. It is plausible that something is not epistemically possible for you if you know it to be false or if its falsity is obviously entailed by something you know. Consequently, a fallibilist can hold that any such assertion would be unacceptable because it is false (Stanley 2005): to assert "I might be mistaken about p" is to assert that it is *epistemically possible* that p is false, which contradicts the claim that one knows that p. Moreover, fallibilism does not in fact require or license one to assert, "I might be mistaken," whenever one knows that p on the basis of a fallible position. A well-designed fallibilist theory will allow that one is sometimes entitled to assert "I know that p." If whenever one asserts, "I know that p," one *cannot* (on pain of contradiction) assert "I might (epistemic possibility) be mistaken," a fallibilist theory can thus avoid any difficulties on this score. It is true that on any fallibilist view one will be entitled to assert, "I know that p, but my justification/evidence/epistemic position regarding p does not entail or necessitate the truth of p." However, such an assertion is not equivalent to "I know that p, but it's epistemically possible that I'm mistaken," and any objection to it will merely reassert a prior commitment to an infallibilist view of knowledge.

A number of writers have recently argued that fallibilist accounts of knowledge are inevitably subject to Gettier-type counterexamples (Owens 2000, Zagzebski 1994, Neta 2002, McDowell 1995, 2002). According to these writers, the failing in Gettier-type cases is that the person's belief is true only by accident or luck (relative to his or her evidence, justification, or epistemic position more broadly construed). The charge, then, is that fallibilist views can't preclude cases in which a person satisfies whatever requirements the fallibilist might impose but the person's belief is nonetheless true only

by accident or luck: just imagine a case in which despite the person's evidence, justification, etc., the belief is false, and then modify the case so that good fortune intervenes to render the belief true.

This argument fails. A fallibilist can avoid the objection by specifying that one doesn't have knowledge unless something like the following condition is met: if one's belief is true, its truth isn't an accident or lucky (relative to one's evidence, justification, or epistemic position). Even if a belief's truth is not entailed by one's evidence, justification, or epistemic position, it doesn't follow that its truth is an accident or lucky (relative to one's evidence, justification, or epistemic position). It is therefore possible to satisfy such a requirement, possess a fallible justification that is adequate for knowledge, and yet fail to have a true belief. But a view which incorporates some such requirement will not be vulnerable to Gettier-type counterexamples (Ryan 1996, Howard-Snyder & Feit 2003).

A closely related worry is that on a fallibilist view, one's possession of a true belief will *always* be a matter of accident or good luck relative to one's evidence, justification, or epistemic position (McDowell 1995, 2002). If having a belief which meets the fallibilist requirements is not inviolably connected with having a true belief, then it would seem that an element of luck intervenes in every case: either bad luck yielding a false belief (unless that bad luck is in turn counteracted by further good luck) or good luck yielding a true belief. However, this is not so. To put the point in the most general terms, the absence of bad luck is not always itself a matter of luck. In the ordinary run of things, it's not a lucky accident that my belief that my yard has moles – based, as it is, on the sorry state of my lawn – is true. However, my evidence, justification, or epistemic position more broadly construed is still plausibly held not to entail or necessitate the truth of that belief. It is a mistake to think that the fallibilist must understand the relation between one's epistemic position and the truth of one's belief as resulting from the operation of a chance mechanism.

Fallibilism can be characterized as the idea that in some cases in which a person knows that p, there is something about the person and his or her circumstances which does not entail the truth of p but which, in combination with the truth of p, makes it the case that the person knows that p. Understood in this way, fallibilism assumes that in some cases, a person's knowing that p is a complex state or state of affairs involving the person, his or her belief, his or her circumstances, and the fact (or true proposition) *p*. Timothy Williamson (2000) has recently challenged this conception of knowledge, arguing that knowledge that p is a non-analyzable factive mental state – a mental state that entails the truth of p. If this view is right, then formulations like the above characterization of fallibilism will fail. However, an approach like Williamson's allows for the possibility that even though knowledge that p is a non-analyzable factive mental state, there are necessary conditions for being in this state. One such condition could be that one have adequate evidence for p, and it could be allowed that this condition can be met even if one's evidence does not entail or necessitate the truth of p. Consequently, even a Williamson-style approach to knowledge can be developed in a way that is reasonably regarded as fallibilist. Though such an approach has some benefits, Williamson's arguments for it are inconclusive (Leite, 2005).

Lotteries and cases of statistical reasoning pose particular problems for fallibilism. The challenge here is to explain how one can ever have knowledge through

statistical reasoning though one can't, it seems, know on merely statistical grounds that one's ticket has lost in a fair lottery (Vogel 1990, Cohen 1988, DeRose 1996, Hawthorne 2004, Harman and Sherman 2004). These problems are compounded for fallibilist views which accept epistemic closure (Vogel 1990, Hawthorne 2004). Solving these problems is arguably the most important challenge currently facing fallibilism's detailed development.

References:

Austin, John (1961). "Other Minds," in *Philosophical Papers*, eds. J. O. Urmson and J.
     G. Warnock, Oxford: Oxford University Press.
BonJour, Laurence (1997). *In Defense of Pure Reason*, Cambridge University Press.
Casullo, Albert (2003). *A Priori Justification*, Oxford University Press.
Cohen, Stewart (1988). "How to Be a Fallibilist," *Philosophical Perspectives* 2:
     Epistemology, J. Tomberlin, ed., Atascadero, CA: Ridgeview Publishing Co.
DeRose, Keith (1996). "Knowledge, Assertion, and Lotteries," *Australasian Journal of
     Philosophy*, 74:4, 568 – 580.
Dretske, Fred (1981). "The Pragmatic Dimension of Knowledge," *Philosophical Studies*,
     40, 363 – 78.
(1970). "Epistemic Operators," *Journal of Philosophy*, 67, 1007 – 23.
Feldman, Richard (1981). "Fallibilism and Knowing that One Knows," *The
     Philosophical Review*, 90: 2, 266 – 282.
Haack, Susan (1979). "Fallibilism and Necessity," *Synthese*, 41, 37-63.
Harman, Gilbert and Brett Sherman (2004). "Knowledge, Assumptions, Lotteries,"
     Philosophical Issues 14 (a supplement to *Nous*) E. Sosa and E. Villanueva (eds.),
     Boston, MA: Blackwell Publishing, pp. 492 – 500.
Hawthorne, John (2004). *Knowledge and Lotteries*, Oxford University Press.
Hetherington, Stephen (1999). "Knowing Failably," *Journal of Philosophy*, 96, 565 – 87.
Howard-Snyder, Daniel; Howard-Snyder, Frances; Feit, Neil (2003). "Infallibilism and
     Gettier's Legacy," *Philosophy and Phenomenological Research*, 66:2, 304 – 327.
Kaplan, Mark (2006). "If you know, you can't be wrong," in S. Hetherington (ed.),
     Epistemology Futures, Oxford University Press, 180 –98.
Leite, Adam (2005). "On Williamson's Arguments that Knowledge is a Mental State,"
     Ratio (new series), vol. XVIII no. 2, pp. 165 – 75.
(2004). "Is Fallibility an Epistemological Shortcoming?" *The Philosophical Quarterly*,
     54:215, 232 – 251.
Levi, Isaac (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Creedal
     Probability and Chance*, Cambridge, Mass.: MIT Press.
Lewis, David (1996). "Elusive Knowledge," *Australasian Journal of Philosophy*, 74, pp.
     649 – 67.
McDowell, John (2002). "Knowledge and the Internal Revisited", *Philosophy and
     Phenomenological Research*, 64, 97-105.
     (1995). "Knowledge and the Internal," Philosophy and Phenomenological
     Research, 55(4):877-893
     (1994). "Knowledge by Hearsay," in Bimal Krishna Matilal and Arindam
     Chakrabarti, eds., *Knowing from Words: Western and Indian Philosophical*

     Analysis of Understanding and Testimony, pp. 195-224. Synthese Library, 230. Dordrecht, Holland & Boston, Mass.: Kluwer Academic, 1994.

Neta, Ram (2002). "S knows that P," *Nous*, 36:4, 663–681.

Nozick, Robert (1981). *Philosophical Explanations*, Harvard University Press.

Owens, David (2000). *Reason Without Freedom*, Routledge.

Reed, Baron. "How to Think About Fallibilism," *Philosophical Studies*, 107, 143 – 57.

Ryan, Sharon (1996). "Does Warrant Entail Truth?" *Philosophy and Phenomenological Research* 56, 183 – 92.

Rysiew, Patrick (2001). "The Context-Sensitivity of Knowledge Attributions," *Nous*, 35, 477 – 514.

Stanley, Jason (2005). "Fallibilism and Concessive Knowledge Attributions," *Analysis* 65:2, 126 – 31.

Vogel, Jonathan (1990). "Are There Counterexamples to the Closure Principle?" in M. Roth and G. Ross (eds.), *Doubting: contemporary perspectives on skepticism*, Kluwer Academic Publishers, 13 – 27.

Williamson, Timothy (2000). *Knowledge and Its Limits*, Oxford University Press.

Zagzebski, Linda (1994). "The Inescapability of Gettier Problems," *Philosophical Quarterly*, 44:174, 65-73.