PHILOSOPHICAL PERSPECTIVES

Philosophical Perspectives, 29, Epistemology, 2015

REBUTTING AND UNDERCUTTING IN MATHEMATICS

Kenny Easwaran
Texas A&M University

1. Mathematical Publishing gives Defeasible Reasons

One of the important tasks for epistemology is to understand the structure of reasons for belief. Importantly, most reasons for belief in general are defeasible.[1] As noted by John Pollock, among others, there are at least two important ways a reason can be defeated. If r_1 is a reason to believe p, it can be *rebutted* if one gets r_2, which is a reason to believe $\neg p$. (To decide what to believe in this case, we may need a notion of "strength" for reasons.) But r_1 can also be defeated by *undercutting*, when one comes to have reason to believe that r_1 is not actually a reason to believe p.

As an example, Emma says she saw Liam in San Antonio on Friday. This gives us reason to believe that Liam was in San Antonio. If Olivia says she saw Liam in Amarillo on Friday, this *rebuts* the previous testimony. If instead Noah told us that Emma is unreliable and often lies about who she saw while traveling, this would *undercut* Emma's testimony. In both cases, we may end up suspending judgment as to Liam's whereabouts on Friday, though in the former case it is likely to matter how reliable we think Emma and Olivia are, while in the latter case all that seems to matter is that we think Noah meets some threshold of reliability. The structure of rebutting and undercutting can get quite complex when very many reasons are involved.

Mathematical reasoning is often taken to be different. However, we should expect it, as a type of human reasoning, to have defeasible features as well. Some might claim that mathematical reasoning just is logical entailment, and logical entailment is indefeasible. After all, defeasible reasoning is non-monotonic, in that adding additional premises can often defeat an inference, while classical logical entailment is monotonic. If the axioms of Peano arithmetic entail that there are no positive integer solutions to $x^n + y^n = z^n$ with $n > 2$, then there is no further premise that can be added to these axioms that can defeat this entailment. However, as Gilbert Harman (1986) has pointed out, inference is not the same thing as implication; reasoning is *informed* by logical entailment but often proceeds in other ways as well. Although the solution to a homework assignment in geometry or logic might take the form of a step-by-step sequence of formulas that follow from each other by syntactically valid rules, this is not the general form of a mathematical proof, and it is not what mathematicians produce in their reasoning.

One of the strongest defenses of the role of logical entailment in mathematical epistemology is given by Azzouni (2004). However, even he notes that the proofs produced by working mathematicians differ significantly from logical entailments, in ways that depend on the knowledge of presumed readers. He notes that "the day-to-day practice of mathematicians isn't to actually *execute* such derivations, but only to *indicate*, to themselves or to others in their profession, such derivations." (p. 95) Although the derivations themselves may not be defeasible, the indication of them provided in a piece of mathematical reasoning surely is.

The result of mathematical reasoning is belief in a mathematical proposition, but it is not certainty. Mason may be able to find the set of solutions of a given second order differential equation. But if he knows Ava is a much better mathematician than him, and if Ava tells him that he made a mistake, he is likely to give up his belief. There are some mathematical propositions that are believed so strongly that it is very hard to see what might defeat one's belief — for instance, the belief that for every integer there is a larger one, or (for more mathematically sophisticated people) the belief that there are infinitely many primes, or the belief that every continuous function of a real variable that takes both positive and negative values must have a zero. But these examples are no more problematic for the claim that mathematical reasoning is generally defeasible than similarly certain empirical propositions are for the claim that empirical reasoning is generally defeasible. I have a hard time imagining what evidence would actually defeat my belief that there is a table in front of me right now, or that Texas is more than ten miles wide. It doesn't matter if there are *some* mathematical beliefs that are as practically certain as these empirical beliefs, or even more so. For *most* mathematical beliefs that are of enough interest to be worth publishing, the reasons we can cite for believing them are defeasible, like the reasons we can cite in scientific publishing.

The phenomenal character of mathematical discovery is that of a defeasible process of inference. This is especially clear when we consider the case of a student struggling with his math homework. But it is also true in the case of mathematicians making discoveries. In doing her work, a mathematician will make a conjecture, then wrestle with the ideas until she has a proof sketch. She will explain the ideas to a friend, partly because the friend is curious, but also partly to get feedback on whether there is an obvious mistake. If it seems to work, she will write up the proof in a manuscript (probably fixing some errors in reasoning as the details are expanded), which is circulated to other mathematicians for verification. Finally, she will submit the revised manuscript to a journal, where referees will check it more thoroughly, and if no mistakes are found, it will be published. At each stage of the process, the mathematician's confidence in the result will increase, but at no point in this process is it absolutely certain.[2] Every mathematician knows of many examples in which false results have made it through this whole process, and one can feel just as confident of something that turns out to be false as of something that turns out to be true.[3] The feeling of discovery is the same in either case, and it is a feeling of confidence, but no more like perfect certainty than the feeling one gets with many sorts of empirical discoveries.

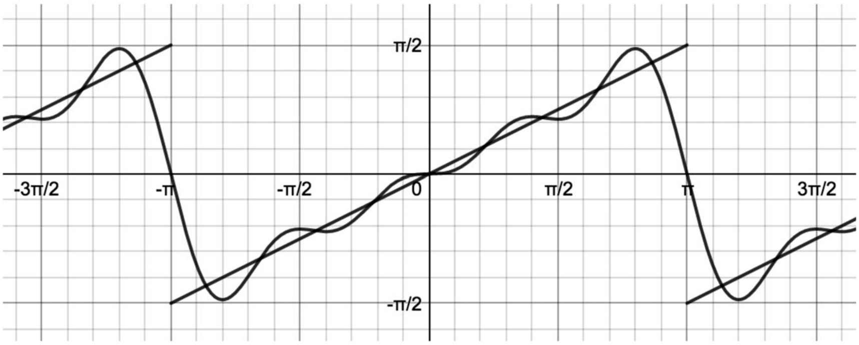
What is published in a mathematical paper is not a set of infallible steps from a basic set that constitute a formally valid proof. The question of what mathematicians actually want with a published proof has no simple and clear answer. (Fallis, 2002, 2003) I claim that thinking about the structure of defeat can help us make this more precise. Although published proofs are not required to be so detailed that they are indefeasible (as the solutions to simple assignments in geometry and logic might be), they are required to have enough detail to impose *some* strong conditions on any potential defeater. In particular I will argue that among other requirements of intelligibility and comprehensibility, there is a requirement about level of detail that allows one to convert rebutting defeat to undercutting defeat. But before I argue for this, I will examine some instances from the history of mathematics in which published proofs were defeated, both by rebutting and undercutting defeaters.

2. Rebutting Defeat in Mathematics

Rebutting defeat exists in many places in the mathematical literature. R.B. Kershner (1968) claimed that all pentagons that can tile the plane are of one of 8 types. Martin Gardner wrote about this paper in his magazine column “Mathematical Games”, and a few years later, Doris Schattschneider noted that the situation had drastically changed. One reader had sent Gardner an example of a ninth type of pentagon that could tile the plane, and amateur mathematician Marjorie Rice had systematically investigated and found three more types. (Schattschneider, 1978) Kershner’s claim had been successfully rebutted by new discoveries.[4]

For a more famous example, consider an early episode in the development of the theory of continuous functions and series. In his 1821 textbook *Cours d’analyse*, Cauchy proved that the limit of a series of continuous functions must always be continuous. His proof was approximately the following. Let the s_n be the sequence of continuous functions, and let s be the function that is its limit. Because $s_n \to s$, we have $s(x+\alpha) - s_n(x+\alpha) \to 0$ as $n \to \infty$, and $s_n(x) - s(x) \to 0$ as $n \to \infty$. Because each s_n is continuous, we have $s_n(x+\alpha) - s_n(x) \to 0$ as $\alpha \to 0$. Putting this all together, we have $s(x+\alpha) - s(x) = [s(x+\alpha) - s_n(x+\alpha)] + [s_n(x+\alpha) - s_n(x)] + [s_n(x) - s(x)]$, and as $\alpha \to 0$ and $n \to \infty$, this sum must also go to 0. Thus, s is continuous at x. (Sørensen, 2005)

However, in a footnote to one of his papers, Abel noted that “this theorem admits of exceptions”. In particular, from his study of Fourier series, Abel noted that the discontinuous sawtooth wave is the sum of the series $\sum_{i=1}^{\infty}(-1)^{n+1}\frac{\sin nx}{n}$, which is a series of continuous functions. (The diagram below includes the sawtooth wave and the fourth approximation in the Fourier series.) Thus, this famous theorem was rebutted. I will discuss this example further in section 4.



3. Undercutting Defeat in Mathematics

Undercutting defeat also exists in the mathematical literature. One classic example is the history of the Four Color Theorem. I describe the outline of the situation here, and provide fuller details in the appendix. The theorem states that every ordinary map can be colored with just four colors in such a way that no two adjacent regions have the same color. Mathematicians phrase it more precisely by representing each region by a vertex in a planar graph, and saying that a map is "ordinary" if we can connect the vertices with edges representing adjacency of regions, in such a way that no two edges cross. (This is meant to rule out cases in which more complex types of adjacency arise from regions that are spatially disconnected, or have complicated infinite boundaries like the example due to Hudson (2003).) The precise statement of the Four Color Theorem is then that in any planar graph where no two edges cross, it is possible to color the vertices with just four colors in such a way that no edge connects two vertices of the same color.

Alfred Kempe published a proof of the Four Color Theorem in 1879, which was shown to have a flaw by Percy Heawood in 1890. The basic outline of Kempe's proof begins by considering a hypothetical counterexample to the Four Color Theorem of minimal size. Because the counterexample is minimal, it must be possible to color the graph with four colors if a single vertex is removed. Kempe then showed that any planar graph must contain a vertex with not too many neighbors. Removing this vertex yields a smaller graph that has a coloring with just four colors. Kempe then considered all the possibilities for how the removed vertex must be related to the vertices in the smaller graph. He showed that there are only a few possibilities, and then gave an algorithm for each possibility to show how to transform the coloring of the smaller graph into a proper coloring of the larger one. Thus, there could be no minimal counterexample to the Four Color Theorem, and thus no counterexample at all.

However, Heawood showed that one of the possible configurations that Kempe considered actually had a special case in which his algorithm would fail. For particular instances of Heawood's configuration, one can find alternate colorings that in fact only use four colors. Thus, Heawood did not find a counterexample to the theorem as a whole. However, he showed that Kempe's proof did not in fact establish the theorem for all cases, and thus undercut the justification that mathematicians had for believing in it. However, a slight modification of Kempe's proof suffices to prove the Five Color Theorem, and it also proves a restricted version of the Four Color Theorem for graphs not containing Heawood's configuration. The full Four Color Theorem was later proved by Wolfgang Haken and Kenneth Appel in 1976, who conducted a more detailed analysis of the configurations that Kempe and Heawood had drawn attention to, and showed (using a computer) that every one of the thousands of special cases of it could in fact be colored with just four colors. So the theorem was eventually shown to be true, even though the original proof was undercut after standing for a decade.

4. Converting Rebutters into Undercutters

Note the difference between Heawood's counterexample to Kempe, and Abel's counterexample to Cauchy. Abel's case is one that shows Cauchy's theorem to be false, while Heawood's case just shows that Kempe's proof can't work. Abel's counterexample is a rebutter to Cauchy's theorem while Heawood's counterexample merely undercuts Kempe's proof.

Interestingly, although Abel's counterexample doesn't by itself show where Cauchy's proof went wrong, we can use it for this purpose. When s is the sawtooth wave, and s_n is the nth sinusoidal approximation to it, we can see that although each $s_n(x)$ is continuous at $x = \pi$, $s(x)$ itself is not continuous at that point. Recall that we "showed" that $s(x+\alpha) - s(x)$ was small by noting that it was the sum of three terms that were themselves small, namely $s(x+\alpha) - s_n(x+\alpha)$, $s_n(x+\alpha) - s_n(x)$, and $s_n(x) - s(x)$. The first and last terms can be made arbitrarily small by choosing n large enough, and the middle term can be made arbitrarily small by choosing α small enough.

The problem with the proof is that these two choices cannot be made at the same time. Cauchy's argument had a quantifier scope ambiguity. For every α there is an n such that $s(x+\alpha) - s_n(x+\alpha)$ and $s_n(x) - s(x)$ are small, and for every n, there is A such that $s_n(x+\alpha) - s_n(x)$ is small whenever $\alpha < A$. But when $x = \pi$ (right at one of the sudden jumps), there are no n and $\alpha > 0$ that simultaneously make all three terms small. By looking at the diagram above (where $n = 4$) we can see that $s_n(x) - s(x) = 0$, but that $s(x+\alpha) - s_n(x+\alpha)$ is small only if $|\alpha| > \pi/5$, while $s_n(x+\alpha) - s_n(x)$ is small only if $|\alpha| < \pi/10$. The former range expands as n gets larger, while the latter range contracts, so there is no way to simultaneously make all three terms small when $x = \pi$. Thus, proper consideration of the rebutting defeater has produced an undercutter for Cauchy's argument. We can use the counterexample to find the step in the proof that was incorrect.

I claim that this is an important feature of most published mathematical argument. There is of course a primary expectation that any published mathematical proof actually be correct. However, since mathematicians only have fallible access to this feature of a proof, there is a secondary requirement that is easier to check, which should apply even if the primary requirement fails. This secondary requirement is that a published proof should contain enough detail that any rebutting defeater can be converted into an undercutting defeater. That is, if (hypothetically) someone were to find a counterexample, one would be able to trace the counterexample through the proof and find out which particular step failed. I will call a defeasible reason for belief "convertible" when it has this sort of structure.

5. Convertibility and "Lemma Incorporation"

This notion of "convertibility" is related to some ideas discussed by Imre Lakatos in his (1976). In this book Lakatos discusses the method of "proofs and refutations" which he says is behind the development of mature and successful mathematics. He suggests that in considering some domain of mathematical objects, one might form a conjecture, and create various "proofs" of this conjecture, while also finding various counterexamples. He proposes the following rules for refining the conjecture into a mature theorem about a related domain of objects:

> Rule I. If you have a conjecture, set out to prove it and to refute it. Inspect the proof carefully to prepare a list of non-trivial lemmas (proof-analysis); find counterexamples both to the conjecture (global counterexamples) and to the suspect lemmas (local counterexamples).
>
> Rule 2. If you have a global counterexample discard your conjecture, add to your proof-analysis a suitable lemma that will be refuted by it, and replace the discarded conjecture by an improved one that incorporates that lemma as a condition. Do not allow a refutation to be dismissed as a monster. Make all 'hidden lemmas' explicit.
>
> Rule 3. If you have a local counterexample, check to see whether it is not also a global counterexample. If it is, you can easily apply Rule 2.[5]

In other parts of the book, Lakatos goes into greater detail about some of the unproductive responses to counterexamples that he rejects: "monster barring" is the act of claiming that a particular counterexample never was a member of the domain to begin with, because it was a "monster"; "exception barring" is the replacement of the domain by a smaller domain that happens to exclude the counterexamples. (The significant difference between these two methods is that monster barring is an attempt to claim that the original domain was never meant to include objects of the pathological sort, while exception barring allows that the original domain did include them and instead changes the conjecture to one about a smaller domain.) Lakatos' general recommendation replaces these two methods by a method he calls "lemma incorporation" on which one shrinks the domain by noting the specific restrictions needed for the proofs of the lemmas. The proofs and the refutations together develop the content of the theorem, rather than just one or the other alone.

In my terminology, rebutting defeaters for mathematical claims usually take the form of global counterexamples, while undercutting defeaters usually take the form of local counterexamples (though I will discuss some other cases later). Lakatos notes that any case in the domain will be a global counterexample, a local counterexample, both, or neither. An ideally developed theorem is one in which no case in the domain is a counterexample of either type. When all counterexamples are both global and local, we can use the method of lemma incorporation to come up with an alternate definition of the domain to improve the conjecture into a theorem of this ideal sort.[6] But when there are global counterexamples that are not local, or local counterexamples that are not global, then the proof itself needs some improvement. If there are local counterexamples that are not global (as in Kempe's proof of the Four Color Theorem) then the lemmas are not all true over the full domain, and so the proof can't prove the full theorem (it could prove a version of the theorem with a suitably restricted domain, but one will need a different proof with new lemmas to extend it to the cases that are local but not global counterexamples). If there are global counterexamples that are not local, then the proof was not convertible in my sense — the analysis of the proof has not yet found all the important lemmas. My requirement of convertibility lines up with Lakatos' ideal of the proper development of a theorem.

Looking back at Cauchy's theorem, we see that the Fourier series for the sawtooth wave is a global counterexample (it is a sequence of continuous functions that converges to a discontinuous function). The main lemma fails for this case because of the quantifier scope ambiguity. Thus, this series is also a local counterexample. However, consideration of this lemma enables us to propose a slightly modified domain of applicability for the theorem. Instead of applying to *all* sequences of continuous functions that converge to a limit, it applies whenever the functions converge in a way that allows us to reverse the scope of the quantifiers.

We say that s_n converge *pointwise* to s iff for every ϵ it is the case that $\forall x \exists N \forall n((n > N) \rightarrow |s(x) - s_n(x)| < \epsilon)$. We say that s_n converge *uniformly* to s iff for every ϵ it is the case that $\exists N \forall x \forall n((n > N) \rightarrow |s(x) - s_n(x)| < \epsilon)$. Consideration of the counterexample reveals that Cauchy's original theorem was phrased in terms of pointwise covergence, but the lemma only works in cases of uniform convergence. By incorporating the lemma, modern mathematicians state the relevant theorem as saying that if the s_n are a sequence of continuous functions that converge uniformly to s, then s is continuous.

Proof: If the s_n converge uniformly to s, then we can choose N big enough to make $s(x+\alpha) - s_n(x+\alpha)$ and $s_n(x) - s(x)$ small for *all* x and α, and then continuity of s_n at x ensures that values of s can't differ too much when α is small.

This definition of "uniform convergence" is a definition that emerges from consideration of the proof, by Lakatos' method of lemma incorporation. This was possible because Cauchy's proof had enough detail for the global counterexample to generate a local counterexample. The convertibility of the proof allowed it to enable this development, even though the proof ended up failing at its original purpose. Mathematicians can't insist on infallible proofs, but they can insist that fallible proofs have this particular kind of fallibility. Any potential counterexample should yield an identification of the false lemma, which enables a replacement of the theorem by one that holds over a slightly different domain, with basically the same proof.

Cases of undercutting defeat without rebutting defeat yield a slightly different analysis. Heawood's challenge to Kempe's proof of the Four Color Theorem (discussed in the appendix) produced a local counterexample that is not global. Kempe's proof (slightly modified) is sufficient to prove the Five Color Theorem, and it proves the Four Color Theorem for a domain of graphs that don't include Heawood's cases as subgraphs. The falsity of one lemma was enough to suggest slightly different theorems (incorporating this lemma or a closely related one) that held with basically the same proof. But the full Four Color Theorem needed new lemmas and new methods, eventually involving computers to prove some of the sub-cases. In cases where there is a rebutter, we don't need a substantially new proof, because the original theorem is not itself true, and incorporating the lemma from the original proof already gives the best theorem in the vicinity. But in the case of undercutters without rebutters one might look for a new proof that covers the cases that are local but not global counterexamples.

Not all undercutting defeat in the absence of rebutting defeat works like this. Sometimes a lemma can be recognized as incorrect without actually providing a local counterexample. This happened in the case of Paris (1994)'s challenge to Cox's Theorem. Cox (1946) claimed to prove that if there is a numerical "believability" function $B(\phi|\psi)$ taking sentences ϕ and ψ in a logical language to values between 0 and 1, and if this function satisfies two further functional constraints, then B can be represented as a probability function. His constraints are that there must be continuous auxiliary functions F and G such that $B(\phi\&\psi|\theta) = F(B(\psi|\theta), B(\psi|\psi\&\theta))$ and $B(\phi|\theta) = G(B(\neg\phi|\theta))$, with F increasing in both arguments and G decreasing. The details of his proof are not important, but Paris noted an important step that Cox made without argument. Cox showed that if $x = B(\chi|\phi\&\psi\&\theta)$, $y = B(\phi|\psi\&\theta)$, and $z = B(\psi|\theta)$, then $F(F(x, y), z) = F(x, F(y, z))$. However, in the next line, he used the claim that this functional equation holds for F, for *any* values $0 \leq x, y, z \leq 1$. Paris noted that since F is assumed to be continuous, it suffices to add a further assumption that B actually take on sufficiently many values that are dense in the unit interval, but this requires a slight modification of the statement of Cox's theorem.

Interestingly, Paris's undercutting defeat for Cox's original theorem did not consist of either a local or a global counterexample — he just noted that Cox's assumption was not justified. The first actual counterexample was given by Halpern (1999a) (which also produced a counterexample to a related theorem of Terence Fine). This counterexample was both a global counterexample (it was a function B satisfying Cox's original assumptions that is not representable as a probability function) and a local counterexample (the function F satisfies associativity over a subset of the interval, but not the whole interval). Had Halpern's counterexample been found first, it would have led to an analysis that turned rebutting defeat into undercutting defeat, but in this case the lemma was recognized as invalid before a specific counterexample was found. Halpern (1999b) gives a fuller analysis of the theorem and its lemmas to show several specific domains over which a version of the theorem is valid. Lemma incorporation proceeded as recommended by Lakatos.

6. Convertibility and Transferability

In my (2009) I characterized arguments as “transferable” if they present a sequence of reasons that an expert will find compelling on mere consideration, with no essential dependence on the authority or trustworthiness of the author. Knowledge that the reader gains from a transferable proof is autonomous, and has no essential dependence on testimony.[7] However, I did not there give much of an explanation for *why* transferability would be a requirement on published mathematical proofs. As I noted, non-transferable proofs may be just as reliable at generating truths and giving any other individual epistemic benefit. The benefit is apparently to the community, in that each member can have a sort of autonomy if she so desires — she can work through all of the justifications just by following the literature, without having to trust prior authors.

However, as Elizabeth Fricker puts it (2006), one of the virtues of epistemic autonomy is that in relying on testimony “one will lack the characteristic sensitivity to defeating evidence, should it come along, which is usually taken to be a hallmark of belief which amounts to knowledge.” This explanation of the value of autonomy suggests that consideration of the structure of defeat will help illuminate the importance of testimony and autonomy.

There are two ways to understand the structure of reasons that the hearer has in a case of testimony, roughly corresponding to the positions known as “reductionism” and “non-reductionism”. (Adler, 2012) The reductionist view suggests that the reasons for belief are that one has heard the testimony, and that one has independent reason to trust the speaker. (This is roughly how I described the cases of testimony as to Liam’s whereabouts on Friday.) The fact that a person who is believed to be reliable has asserted something gives us a reason to believe the content of what they have asserted. The non-reductionist view suggests that although beliefs about the speaker can defeat the testimony, they are not themselves part of the reason for belief of the content of what is said. One way to think of this is that the testimony somehow provides an encapsulated version of the reasons that were available to the testifier. When a claim is asserted by a person who in fact has reason to believe it, then the hearer in some sense has that very same reason.

On the reductionist view, the fact of testimony creates new reasons for the hearer, but these reasons are very different from the ones that the testifier had. On the non-reductionist view, the testimony cannot create new reasons, but can only give indirect access to existing reasons. Interestingly, if the testifier has insufficient reason to believe the claim (as when the claim is a lie, or when the testifier was irrational in coming to believe it), then the hearer also has insufficient reason to believe the claim. Thus, on this view, the hearer may not know whether she has sufficient reason for belief, and what kind of reasons she does or does not have.

Importantly, on both of these views, a reliance on testimony eliminates convertibility. For the reductionist, the testimony provides a new reason, but it is a reason of a special sort based on testimony. The reason that the hearer has for believing a claim is whatever reason she has for trusting the testifier, together with the fact that the testifier has made the claim. If the hearer gets a rebuttal, then no part of this reason is necessarily undercut. On learning that a claim made by your very reliable friend or teacher has turned out to be false, one does not need to either give up the belief that the friend or teacher is generally reliable, or the belief that the friend or teacher made the claim. Both parts of the reason can survive intact despite a rebuttal. Thus, the rebuttal does not generally create an undercutter. (In special cases it may be able to, as when one only had very minimal evidence for the reliability of the testifier. But it seems that these cases should be fairly uncommon, since one needs to have enough evidence of reliability to trust the person for the initial claim, but not enough that the trust will survive a single instance of falsehood.)

For the non-reductionist, the testimony does not provide a new reason, but convertibility is still lost. Consider Kershner's argument that his 8 pentagons are the only ones that tile the plane. He says, "The proof that the list in Theorems 1 and 2 is complete is extremely laborious and will be given elsewhere." (p. 840) Presumably, this means that Kershner had some mathematical argument in mind that had given him reason to believe that these 8 pentagons are the only ones that tile the plane. On seeing the new pentagons produced by Marjorie Rice, Kershner could probably work through his argument and figure out which step relied on a mistake.[8] But this is certainly not true for anyone who believed the claim on the basis of Kershner's testimony. Although the hearer may in some sense have a belief that is supported by reasons that were undercut by this counterexample, the hearer is in no way able to convert this rebutter into the undercutter.[9]

Thus, we don't need to settle the debate between the reductionist and the non-reductionist about testimony to say that an essential reliance on testimony can interfere with convertibility of one's reasons for belief. And these ideas are related to the way that a reliance on testimony blocks transferability. In a transferable proof, the reader has direct access to the reasons that the author had for her belief. But in the case of testimony, the reader either has access to different reasons, or has indirect access to the same reasons.

Regardless of the way that testimony interferes with convertibility, there are certainly other sorts of reasons for belief that are not convertible. For instance, statistical evidence can provide reasons for belief for which a rebutter cannot be converted into an undercutter. Many statistical tests work by observing sufficiently different frequencies of some effect between samples of two populations, such that it would be extremely unlikely for samples to have this difference if there were not some different in frequency in the underlying populations. For instance, one might observe the occurrence of various cancers in people who eat high quantities of some food compared to people who eat low quantities of that food. Some difference in frequency between the population is expected due to chance, but if one observes a difference that has a sufficiently small probability if chance is the only source of variability, then one can get a strong (though defeasible) reason to believe that the food helps cause or prevent cancer.

Schoenfeld and Ioannidis (2012) did a systematic survey of statistical studies for 50 randomly selected ingredients, finding that for most ingredients there had been at least one study investigating its connection with cancer. For almost all ingredients, some statistically significant effect was found, and for many of them there were published studies showing statistically significant effects in opposite directions. The evidence from such a study could theoretically be undercut by finding systematic differences other than diet between the two samples, but in most of these cases it appears that the contradictory observations weren't connected with any associated problems. Some of these studies rebut each other without undercutting the support that they each provide for opposite conclusions.

In my (2009), I argued that statistical arguments in mathematics are rejected because, in published form, they essentially rely on testimony. However, regardless of the role of testimony, the contention here that convertibility is an essential feature for published mathematical proofs gives yet another reason for rejecting these statistical arguments. A requirement of convertibility rules out a reliance on statistical arguments whether or not their social spread relies essentially on testimony.

Transferability was a condition on correct proofs — they should be such that an expert can convince herself of the relevant claims without relying on the author's testimony. Convertibility is a condition on potentially *incorrect* proofs — they should be such that any counterexample can reveal the incorrect step, which can allow us to replace the claimed theorem with a related theorem by the method of lemma incorporation. Transferability rules out a reliance on testimonial or statistical arguments. Convertibility may go farther in ruling out certain inductive or abductive arguments as well. (This may be a way in which mathematics differs from philosophy and other humanities, which seem to allow such arguments, even if I was right that they require transferability in their published arguments.)

Convertibility would not be an appropriate condition for scientific arguments, because it would rule out too much. In mathematics, it can help make sure that fallible proofs contribute to the production of new theorems even if they turn out to be incorrect. Although lemma incorporation doesn't exist in science, undercutting defeat can still play a similar role. Discovering that a statistical sample was not representative of an entire population might allow one to use that sample to draw an inference about a sub-population for which the sample was representative. Scientific practice is set up to maximize the possibility of discovering this sort of undercutting defeat. Scientists are expected to describe their experimental methodology, so that an expert reader could identify these confounding variables and thus modify their understanding of the significance of the observed results. But convertibility is a distinctive feature of mathematical practice.[10]

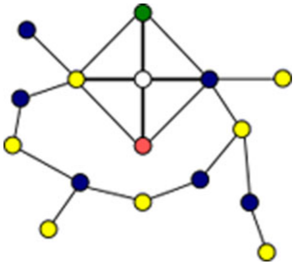
Appendix: Kempe's "Proof" of the Four Color Theorem

This discussion is adapted from lecture notes by Ralph Bravaco and Shai Simonson at Stonehill College available at: http://web.stonehill.edu/compsci/lc/four-color/four-color.htm.

The first step in Kempe's proof is noting that if we can add extra edges to a graph and still color it with just four colors, then the original graph must be colorable with just four colors as well. Thus, we can assume that our minimal counterexample has enough edges that every face in the graph is a triangle. In this case, the number of faces is 2/3 the number of edges (since each face has exactly three edges, and each edge is on exactly two faces), so Euler's famous formula that $V - E + F = 2$ becomes $V - E/3 = 2$, or $E = 3V - 6$. However, if every vertex had at least 6 edges coming out of it, then the number of edges would be at least $3V$ (since each edge comes out of 2 vertices). Therefore, there must be some vertex with at most 5 edges, in order to have $E < 3V$.

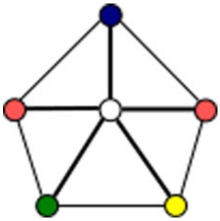
Removing some vertex with at most 5 edges, we have a smaller graph. Thus by minimality of our counterexample, this smaller graph can be colored with four colors. If the vertex that was removed had neighbors using at most three different colors, then it can be colored with the fourth color. Thus, any minimal counterexample must have a vertex with either 4 or 5 neighbors, and in the coloring we get by removing this vertex, these neighbors must use all four different colors. The goal will be to show how to modify the four-coloring of the smaller graph in order to put the removed vertex back in and assign it one of the four colors.

If this vertex has 4 neighbors, then we can proceed as follows. Choose one vertex adjacent to this center and switch it to the color of the vertex opposite this center, say, yellow to blue. If this works, then we can now color the center vertex yellow. This can only fail if the yellow vertex is already adjacent to some blue vertices, in which case we should switch those ones from blue to yellow. Continue following any paths from the vertex, switching blue and yellow vertices. If we eventually run out without reaching the opposite vertex from the center, then we can color the center vertex yellow. But if some "Kempe chain" of alternating blue and yellow vertices extends all the way around, this won't work:

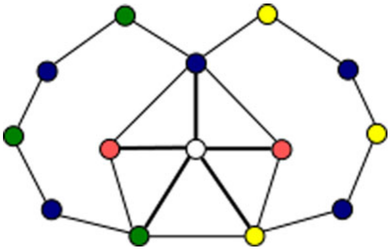


In that case, we consider the other two vertices adjacent to the center, and try switching the red one to green, so we can color the center vertex red. The only way this can fail is if there is a red-green Kempe chain connecting the red vertex to the green one. But a yellow-blue Kempe chain and a red-green Kempe chain can't cross, so this can't happen. Thus, as long as the center vertex has just 4 neighbors, we can four-color the graph with the center vertex just by modification of the four-coloring of the graph without it.

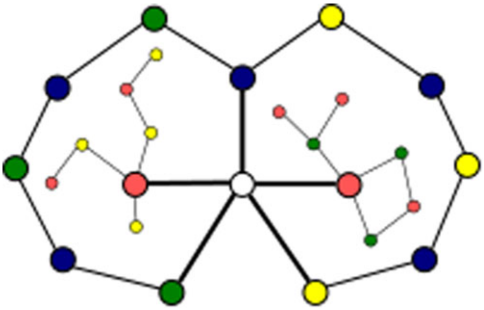
If the center vertex has five neighbors, then two of them must be the same color in the underlying coloring. (If any more than two are the same color, then some color must be unused, and we can just color the center vertex with that color.) Since we have assumed that every face in the graph is a triangle, each neighbor of the center vertex has an edge connecting it to each of the two adjacent neighbors, so the same-color neighbors of this center vertex must not be adjacent to each other. Let's assume that the two same-color neighbors to the left and right are red, that the blue neighbor is above them on one side, and that the yellow and green neighbors are below them on the other side.



We can start by trying to turn the blue neighbor yellow or green. The only way this can fail is if the blue neighbor is connected to the yellow neighbor by a blue-yellow Kempe chain, and is also connected to the green neighbor by a blue-green Kempe chain.



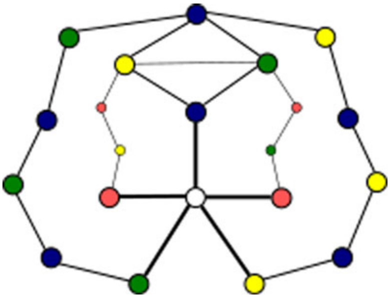
But in this case, we have one more option — if we can turn both red neighbors into other colors, then we can color the center vertex red. So the red vertex surrounded by the blue-green Kempe chain should turn yellow, and the red vertex surrounded by the blue-yellow Kempe chain should turn green. Any red-yellow Kempe chain will be interrupted by the blue-green one before it can hit the red-green Kempe chain coming off the other side, and any red-green Kempe chain will be interrupted by the blue-yellow one before it can hit the red-yellow Kempe chain coming off the other side.



Thus, even this configuration can be given a four-coloring, so no smallest counterexample can exist. At least, if Kempe's proof is to be believed.

The problem is that Heawood noticed that the blue-yellow Kempe chain and the blue-green Kempe chain don't have to go in separate directions, as suggested by the diagram above. Instead, they can cross at a blue vertex. This then allows the red-yellow Kempe chain to come in contact with the red-green Kempe chain:

It's not hard to find a four-coloring of this graph (just change the bottom-right yellow neighbor to blue, and the blue vertex adjacent to it to green, and then the center vertex can become yellow), but Kempe's method doesn't do it, so Heawood's graph undercut Kempe's proof.



Computer-aided proofs of the theorem analyze this case more thoroughly and show that each possible configuration can be colored in some more complex way. But the proof does show that for every planar graph not containing a version of Heawood's configuration, four colors suffice. And even without computer-aided proofs, it's straightforward to modify this argument to show that every planar graph can be colored with *five* colors — the only possible failure would be a vertex with five neighbors all receiving different colors. In this case, the only way to block a modification of the coloring would require a Kempe chain connecting each neighbor with the two opposite neighbors. But this is impossible, because the red-green Kempe chain can't cross the blue-yellow one (or the orange-yellow one, which also can't cross the blue-green one, which itself can't cross the orange-red one, etc.)

Notes

1. I won't worry here about whether reasons are propositions, beliefs, facts, pieces of evidence, or whatever. I also won't worry about what it takes to "have" them. I suspect most of these issues are incidental to the project I'm interested in, and all examples can be described under various answers to these questions.
2. As John Pollock notes,

   > mathematicians tend to be very cautious about accepting the results of complicated proofs. But the proof, if correct, is simply an exercise in deductive reasoning. How is such caution possible? Why doesn't their rational architecture force them to just automatically accept the conclusion? The answer seems to be that they have learned from experience that complicated deductive reasoning is error prone. (Pollock, 1989)

3. In addition to the examples discussed in later sections, see this discussion: http://mathoverflow.net/questions/35468/widely-accepted-mathematical-results-that-were-later-shown-wrong
4. An interesting update to this problem was recently reported in the popular media: http://www.theguardian.com/science/alexs-adventures-in-numberland/2015/aug/10/attack-on-the-pentagon-results-in-discovery-of-new-mathematical-tile
5. Lakatos also suggests a way to widen the domain of applicability of a theorem by a more careful proof analysis instead of just narrowing it in light of counterexamples. This is his "Rule 4. If you have a counterexample which is local but not global try to improve your proof-analysis by replacing the refuted lemma by an unfalsified one."
6. Charlotte Werndl (2009) notes that there may in fact be examples in mathematics where domains should be defined in ways other than by this method of lemma incorporation, but she notes that this method is still an important and quite widely applicable one.
7. Of course, a non-expert reading a transferable proof may not be able to follow every step, and even an expert reading such a proof may not want to expend the mental energy to follow every step, so it may well be that for most readers, there is a step for which reliance on testimony or authority is essential. However, the point is that there is no step such that for most relevant readers, reliance on testimony or authority is essential.
8. Interestingly, other than Kershner's 8 pentagons, all pentagons that are known to tile the plane involve arrangements where the edges of two adjacent pentagons line up exactly along a single edge of another pentagon, and perhaps Kershner's unpublished argument made use of an illicit assumption that this wouldn't

happen. However, neither Kershner nor anyone else has published a proof that these 8 pentagons are an exhaustive list even subject to this extra condition. Conjecturing that a theorem holds subject to this extra condition in the absence of even a proof sketch would be an instance of Lakatos's "exception barring". But if there were a clear proof that could proceed from this extra condition, then this would be an instance of lemma incorporation.

9. Note that Kershner's argument is not a counterexample to my claim that generally, published mathematical arguments must be convertible. Rather than publishing his result in a purely professional journal, he published it in *The American Mathematical Monthly*, a more general interest journal. As their description states:

> The *Monthly*['s] ...readers span a broad spectrum of mathematical interests, and include professional mathematicians as well as students of mathematics at all collegiate levels. ...The *Monthly*'s readers expect a high standard of exposition; they expect articles to inform, stimulate, challenge, enlighten, and even entertain. *Monthly* articles are meant to be read, enjoyed, and discussed, rather than just archived. Articles may be expositions of old or new results, historical or biographical essays, speculations or definitive treatments, broad developments, or explorations of a single application. Novelty and generality are far less important than clarity of exposition and broad appeal. Appropriate figures, diagrams, and photographs are encouraged.

This journal does not mean to publish the official arguments for new mathematical claims, but just to tell readers about interesting ideas that are being worked on.

10. I presented an earlier version of this paper at a workshop in memory of John Pollock at the University of Arizona in 2012, and I would like to thank the participants there for helpful comments. I would also like to thank the participants in the first Texas Epistemology Extravaganza in 2015, where I presented a more recent version, and especially Sinan Dogramaci and Justin Fisher.

References

Adler, J. (2012). Epistemological problems of testimony. *Stanford Encyclopedia of Philosophy* .

Azzouni, J. (2004). The derivation-indicator view of mathematical practice. *Philosophia Mathematica*, 12(2):81–106.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13.

Easwaran, K. (2009). Probabilistic proofs and transferability. *Philosophia Mathematica*, 17(3):341–362.

Fallis, D. (2002). What do mathematicians want? probabilistic proofs and the epistemic goals of mathematicians. *Logique & Analyse*, pages 179–180.

Fallis, D. (2003). Intentional gaps in mathematical proofs. *Synthese*, 134(1-2):45–69.

Fricker, E. (2006). Testimony and epistemic autonomy. In Lackey, J. and Sosa, E., editors, *The Epistemology of Testimony*. Oxford University Press.

Halpern, J. (1999a). A counterexample to theorems of Cox and Fine. *Journal of Artificial Intelligence Research*, 10:76–85.

Halpern, J. (1999b). Cox's theorem revisited. *Journal of Artificial Intelligence Research*, 11:429–435.

Harman, G. (1986). *Change in View: Principles of Reasoning*. MIT Press.

Hudson, H. (2003). Four colors do not suffice. *The American Mathematical Monthly*, 110(5):417–423.

Kershner, R. B. (1968). On paving the plane. *American Mathematical Monthly*, 75:839–844.

Lakatos, I. (1976). *Proofs and Refutations*. Cambridge University Press.

Paris, J. (1994). *The Uncertain Reasoner's Companion*. Cambridge University Press.

Pollock, J. (1989). OSCAR: a general theory of rationality. *Journal of Experimental and Theoretical Artificial Intelligence*, 1(3):209–226.

Schattschneider, D. (1978). Tiling the plane with congruent pentagons. *Mathematics Magazine*, 51(1):29–44.

Schoenfeld, J. and Ioannidis, J. (2012). Is everything we eat associated with cancer? A systematic cookbook review. *The American Journal of Clinical Nutrition*, 97(1):127–134.

Sørensen, H. (2005). Understanding Abel's comment on Cauchy's theorem. *Historia Mathematica*, 32:453–480.

Werndl, C. (2009). Justifying definitions in mathematics — going beyond Lakatos. *Philosophia Mathematica*, 17(3):313–340.