Aha! Trick Questions, Independence, and the Epistemology of Disagreement

Michael Arsenault*    Zachary C. Irving †‡

Published in Thought: A Journal of Philosophy, 2012, Volume 1, Issue 2

Abstract

We present a family of counter-examples to David Christensen’s Independence Criterion, which is central to the epistemology of disagreement. Roughly, Independence requires that, when you assess whether to revise your credence in P upon discovering that someone disagrees with you, you shouldn’t rely on the reasoning that lead you to your initial credence in P. To do so would beg the question against your interlocutor. Our counter-examples involve questions where, in the course of your reasoning, you almost fall for an easy-to-miss trick. We argue that you can use the step in your reasoning where you (barely) caught the trick as evidence that someone of your general competence level (your interlocutor) likely fell for it. Our cases show that it’s permissible to use your reasoning about disputed matters to disregard an interlocutor’s disagreement, so long as that reasoning is embedded in the right sort of explanation of why she finds the disputed conclusion plausible, even though it’s false.

We present a family of counter-examples to David Christensen’s Independence Criterion, which is central to debates in the epistemology of disagreement. The criterion is as follows:

**Independence**

In evaluating the epistemic credentials of another’s expressed belief about P, in order to determine how (or whether) to alter my own belief about P, I should do so in a way that doesn’t rely on the reasoning behind my initial belief about P (Christensen 2011, p. 2)

Our counter-examples involve questions with an easy-to-miss trick that you catch, but have reason to think your peer has not. We argue that in these examples, it’s permissible to use your first-order reasoning to disregard an interlocutor’s disagreement, so long as that reasoning is embedded in the right sort of explanation of the psychological fact that she finds the disputed conclusion plausible, even though it’s false.

*Berkeley, Philosophy Department
†University of Toronto, Philosophy Department
‡First authorship is joint, since all sections were written collaboratively. Send correspondence to zac.irving@utoronto.ca
1 Background

The epistemology of disagreement focuses on whether and how much you should change your credence in a proposition $P$ when you become aware that an epistemic peer has a different credence in $P$. Adapting Kelly's (2005) definition, people count each other as epistemic peers with respect to $P$ iff they take each other to be i) equally familiar with the evidence and arguments that bear on $P$ and ii) equal with respect to general and domain-specific epistemic virtues relevant to reasoning about $P$ (intelligence, expertise, etc.). Intuitively, if you disagree with someone who shares your evidence and relevant epistemic virtues, this gives you reason to believe you made a mistake (and consequently, to reduce your confidence).

Independence (stated above) is meant to explain what's epistemically impermissible about certain question-begging dismissals of the evidence provided by disagreement. For example, it begs the question for a $P$-believer to say, “Well, so-and-so disagrees with me about $P$. But since $P$ is true, she's wrong about $P$. So however reliable she may generally be, I needn't take her disagreement about $P$ as any reason at all to question my belief” (Christensen 2011, p. 2). We agree that responses of this kind beg the question, but argue that Independence fails to capture why.

2 Trick Questions: A Counter-Example

In this section, we identify a family of counter-examples to Independence. To do so, we present a schema that captures the intuitive idea behind these examples and a particular instance of this schema.

---

1 Elga (2007) gives a more general and compact definition of epistemic peer-hood: you count someone as an epistemic peer with respect to $P$ iff you think that, conditional on the two of you disagreeing about $P$ (and the circumstances of disagreement such as whether you are both sober), you are equally likely to be mistaken. Our argument applies, whichever definition you adopt. We work with Kelly's because it helps to motivate an objection we consider later in the paper.
Aha! Trick Questions, Independence, and the Epistemology of Disagreement

**Trick Question Schema (TQS)**

Suppose that you and a friend have a puzzle-solving competition and that you take him to be your epistemic peer with respect to solving puzzles of this sort. You each work through a puzzle (e.g. a math problem) independently, and then compare answers. While you are solving the puzzle you fall for a trick which you later discover. Upon discovering this trick, you think to yourself, “Aha! That was a tempting (but mistaken) way to go!” When you and your friend exchange answers, you find that they’re close, but distinct. In fact, it turns out that your friend has that answer which you think you would’ve gotten had you not corrected your mistake (hereafter, “the decoy answer”).

Intuitively, peer disagreement gives you much less cause to reduce your confidence when the disagreement fits TQS\(^2\) than when it does not. Briefly, the reason is this: since you and your friend are epistemic peers and you almost fell for the trick, it’s likely that he did fall for it, given the way things turned out. To see this, consider an instance of TQS. Suppose that you and your friend have been studying GRE math together for a while. You’ve always gotten roughly the same score, and in particular you’ve always been roughly as likely to get both probability and tricky questions right. As you’re both working through a practice test (at the same time but separately), you come across this question:

**Tricky Primes**

Which is greater?

A. The probability of obtaining an even number on one die and an odd number on the other when two unbiased dice are thrown.

B. The probability of obtaining an even number on one die and a prime number on the other when two unbiased dice are thrown.

C. Neither, because A = B

In Tricky Primes, the decoy answer is C. One might reason that, just as there are three evens and three odds on a die, there are three evens and three primes (recall that 2 is a prime and 1 is not). Therefore, one might reason that A and B are equiprobable: i.e. \( p(A) = p(B) = \frac{(3 \times 3) + (3 \times 3)}{36} = 1/2 \). But in calculating \( p(B) \), we double-counted the combination \( \langle 2, 2 \rangle \), since we counted it once when we thought of 2 as an even and once when we thought of 2 as a prime. To get the correct answer, we must therefore recognize that \( p(B) = \frac{18}{36} - \frac{1}{36} < \frac{18}{36} = p(A) \).

\(^2\)These examples seem closely related to a category of problems that psychologists call “insight problems”. In order to solve a typical insight problem, one must break out of an unproductive way of structuring it (Knoblich et al. 1999). Doing so is associated with the kind of “Aha!” experience we describe above. We commonly face such problems in ordinary life.
Now suppose that, in the course of working through Tricky Primes, you first choose C, but then realize your mistake and switch to A. Upon comparing answers, you find that your friend got C. So the question is whether, and how much, you should reduce your credence in A in response to finding out that he thinks C is correct. Intuitively, you should reduce your confidence only slightly, if at all. And that’s because the step in your first-order reasoning where you caught the trick gives you evidence that the question contains a trick that someone of your ability would likely fall for. You happened to avoid this trick. But since you almost fell for it, your friend (who got the decoy answer) likely did. Thus, you may dismiss (or greatly discount) the evidence you would otherwise get from disagreement. Tricky Primes violates Independence because it’s permissible to use your first-order reasoning (among other considerations) as evidence that you’re in an instance of TQS, which gives you a reason to dismiss the evidence provided by disagreement. We’ll sharpen this argument in the course of responding to objections.

3 Objections

In response to Tricky Primes, our opponents could 1) deny that you use your first order-reasoning, 2) deny that you and your friend are epistemic peers, or 3) cite a different epistemic principle that problematizes Tricky Primes.

To develop 1), our opponents must show that in Tricky Primes you appeal to something other than your first-order reasoning to dismiss your peer. In communication, Christensen has identified one thing to which you might appeal. While doing the problem, you might notice the following pattern in how you reason: you seem to come upon an obvious answer, seem to spot a trick that had mislead you into getting this answer, and then seem to find a better answer that avoids the trick. In other words, you seem to be in an instance of TQS. And in the past, when things have seemed this way to you and when your peer has announced the decoy answer, you’ve been right more often than your peer. If you dismiss your peer on the basis of these seemings, you’re dismissing him on the basis of second-order reflections about the general course your first-order reasoning took, rather than on its mathematical content. Thus, you don’t violate Independence. Here’s another reason to think you satisfy Independence: suppose, before you look

\(^3\)In addition to Christensen, we thank Jonathan Weisberg and an anonymous reviewer for Thought for helping to articulate this objection.
at Tricky Primes, you’re asked, “If you do this problem, and you seem to find yourself in a clear instance of TQS, and your friend announces what seems to you to be the decoy answer, who’d be more likely to be right?” You’d answer that you would. This appears to show that you needn’t rely on your first-order reasoning itself: after all, you can accept this conditional before reasoning at all.

Before developing our reply, let’s start by identifying precisely what differs between Christensen’s interpretation of Tricky Primes and ours. According to us, your justification for thinking there’s a trick in the question is a step in your reasoning: namely, the step where you discovered a trick. Based on this and the fact that your friend got the decoy answer, you dismiss the evidence provided by disagreement. According to Christensen, you dismiss the disagreement on the basis of a higher-order fact about how things seem to you: i.e., your reasoning seems to have taken a characteristically tricky course.

We think our interpretation better captures how you’d intuitively dismiss your peer in Tricky Primes. When your peer announces that she got C, it would be perfectly natural to appeal to the first-order mathematical trick you think she’s fallen for. For example, you might think to yourself, ‘Aha! That sucker double-counted 2!’ You’d feel no rational pressure to retreat to the second-order fact that you seem to be in an instance of TQS. These intuitions aside, even if Christensen is correct that you often appeal only to second-order seemings in cases like Tricky Primes, this doesn’t show that you can’t appeal to your first-order reasoning as additional evidence that your friend made a mistake. Below, we’ll argue that you can. And if we’re right, Independence still fails, since your first-order reasoning can be part of the evidence you use to dismiss your peer.

Imagine that, when your friend announces that his answer is C, it strongly seems that your own reasoning took a tricky course with decoy answer C. If you wanted to provide additional evidence that your friend made a mistake, you probably wouldn’t look for additional evidence about how things seem to you. Rather, you’d want to know whether the way things seem to you is the way they actually are: i.e., whether you’re actually in an instance of TQS with decoy answer C. Prima facie, this is already in tension with Christensen’s thesis that your only reason for dismissing your peer is how things seem to you. Moreover, to provide additional evidence that you’re in an instance of TQS with decoy answer C, it would be natural to double-check your reasoning to ensure that one might come upon C by falling for a trick. You’d want to make sure, for example, that you didn’t initially come to the “decoy answer”
C because of a simple arithmetical error, since you have no reason to believe that your epistemic peer wouldn’t catch such an mistake. To double-check this, you might think the following, “At first, I went about calculating $p(B)$ as follows: add a) the nine combinations where there’s an even number on the first die to b) the nine combinations where there’s a prime number on the first die, which means that $p(B) = \frac{18}{36} = \frac{1}{2}$. But doing this double-counts the combination where 2 appears on both dice: I counted it once when calculating a) and again when calculating b). To get the correct answer, I therefore needed to subtract one combination to get that $p(B) = \frac{17}{36} < p(A)$. Since it takes a subtle arithmetical insight to notice the double-counting, it’s not surprising that my friend missed it (after all, I almost did). This explains why she’s confident in the decoy answer.”

Crucially, your claim that you need to subtract one combination in order to avoid double-counting is both what you present as evidence that one might come upon C by falling for a trick, and part of your first-order reasoning for your conclusion that $p(A) > p(B)$. The step in your reasoning where you catch the trick is thus “Janus Faced”, in virtue of allowing you to see what’s right about the correct answer and why someone would get the decoy answer. Moreover, unlike cases of patent question-begging, this use of your mathematical reasoning appears cogent: you generalized from the specific trick involving double-counting to the conclusion that you’re in an instance of TQS (and thus, that your friend likely made a mistake).

Of course, Christensen could respond that you are begging the question, despite appearances. But since Independence is supposed to capture what’s intuitively wrong about question-begging dismissals of disagreement, it’s a strike against the principle if it clashes with our intuitions about what begs the question. Alternatively, Christensen might respond that we’ve mis-described what’s going on when you double-check your reasoning in the way we dramatized above. Rather than checking your mathematical reasoning to ensure that falling for a trick could lead one to C, you’re merely rehearsing your initial reasoning to determine whether it still seems tricky to you. In other words, you’re trying to determine whether your reasoning reliably produces the relevant sort of seeming in an agent like yourself.

But it’s hard to believe that the ordinary GRE test-taker would use this peculiar strategy to increase his justification for dismissing his peer: namely, the strategy of checking whether he reliably has the same second-order responses to multiple iterations of his first-order reasoning. Moreover, it’s not even clear that
Aha! Trick Questions, Independence, and the Epistemology of Disagreement

this rehearsal strategy would work in the case of Tricky Primes. Generally, once you’ve discovered a trick, you cease to be compelled by the reasoning that originally tricked you\(^4\). Thus, rehearsing your reasoning wouldn’t help you check whether someone would reliably have the second-order response of temptation, followed by the ‘Aha!’ of discovering the mistake in their tempting reasoning. Finally, Christensen’s rehearsal response doesn’t handle cases where a reasoner doesn’t merely rehearse his reasoning, but instead attempts to use his mathematical reasoning as evidence that C is the decoy answer. Once again, Christensen owes us an account of why such a reasoner begs the question when, intuitively, he does not.

We want to conclude with a point about why we structured our response as we did. In a self-reflective agent, the seemings Christensen identifies are like shadows cast by the agent’s first-order reasoning: they’re always present, at least in cases it’s easy to think one’s way into. As such, it’s not obvious that one could decide between our interpretation and Christensen’s by describing a thought-experiment where you use your first-order reasoning to justifiably dismiss a disagreement, but in which the relevant seemings aren’t present. But the fact that these seemings are always present doesn’t show that you always rely upon them to dismiss your friend’s disagreement. On the face of it, it’s plausible that you sometimes appeal to your first-order reasoning in Tricky Primes, and unnatural to say that you’re in fact appealing to how things seem to you in these circumstances. So the balance of arguments fall against Christensen’s objection.

A second objection is that you can dismiss the evidence provided by your friend’s disagreement, not because your friend is an epistemic peer who missed an easy-to-miss trick, but rather because she’s not your peer at all. On Kelly’s definition of peer-hood, someone can fail to be your peer either due to asymmetries in your general epistemic virtues (e.g., intelligence or mathematical ability) or asymmetries in your familiarity with the evidence/arguments. Given the construction of our case, there aren’t significant asymmetries in your general epistemic virtues.

It’s been suggested to us in conversation that, because you discovered the trick in Tricky Primes and your friend did not, you don’t have the same evidence. We think this objection misuses the notion of evidence. In Tricky Primes, it’s natural to contrast your evidence (perhaps the written problem) with your reasoning, including the step where you caught the trick. But let’s concede that you have evidence of

\(^4\)In this respect, feelings of trickiness differ from perceptual illusions, which often persist after you’ve discovered their illusory character.
trickiness that your friend lacks. According to Independence and our definition of epistemic peer-hood, you can’t rely on your first-order reasoning to judge that your friend is less familiar with the evidence, and so is not your peer. But as we’ve argued above, you can rely on your reasoning to infer that your friend likely missed the trick and thus, under this description, is less familiar with the evidence. Therefore, our objector’s re-description still violates Independence.

Lastly, one might object that the reasonable response to disagreement is to remain steadfast until you hear what your peer has to say in support of their position. Until then, you should remain confident in your position. In other words, one could hold that disagreement is epistemically significant only in cases of what Feldman calls “full disclosure” (2006), where you and your peer knowingly share all of your relevant arguments and evidence. But many trick questions are such that, if you were to show your reasoning to an epistemic peer, they would catch the trick as well, and the disagreement would dissolve. There would then be no occasion to use your first-order reasoning to overturn the evidence provided by disagreement, and thus no counter-example to Independence.

While a complete discussion of full disclosure is outside the bounds of this paper, here’s a reason to think it’s overly restrictive. Though it’s often reasonable to engage in full disclosure, in cases like the following, it’s not:

**Math on a Plane** Samuel is getting on a plane to give a math proof at a conference. In the airport he bumps into a colleague who’s working on the same problem but got a different answer. Samuel doesn’t have time to talk through his proof before his flight. Question: should Samuel reduce his confidence and recheck his answer on the plane, even though doing so will come with costs? For example, if he checks his answer, Samuel won’t have time to create a power-point presentation.

Intuitively, Samuel should re-check his answer. And he should do so precisely because discovering that his colleague disagrees with him gives Samuel reason to reduce his confidence. But if full disclosure is correct, then Samuel should suspend judgement on whether to reduce his confidence until he gets back from the conference and has a chance to talk to his colleague. One might respond that Samuel should re-check his answer, but maintain confidence in his conclusions while he does so. But such a view attributes a mental state to Samuel that’s difficult to make sense of: he would be relatively certain that he’s correct, yet feel the need to re-check his answer anyway, despite the substantial costs. Surely a more perspicuous
explanation of Samuel’s (fully rational) decision to recheck his work is that his confidence has decreased.

4 Positive Proposal

Our counter-example doesn’t just defeat Independence, but also puts pressure on Christensen’s characterization of Higher Order Evidence (HOE). In the interests of remaining neutral, we’ll characterize HOE in terms of some of Christensen’s examples. Suppose I’m confident that I’ve answered a puzzle correctly, but also know I’ve taken a drug. In 80% of cases, the drug impugns a person’s puzzle-solving ability even though they remain very confident in their answers. My awareness that I took the drug is HOE that I made a mistake. Another example of HOE is peer disagreement, since it typically gives one evidence that one made a mistake (Christensen 2010, pp. 2-4).

For Christensen, an important general feature of HOE is that it’s subject to a bracketing principle closely related to Independence: “[i]n accounting for…HOE…, I must, in some sense,…put aside or bracket my original reasons for my answer” (ibid., p. 195). For instance, I cannot appeal to the apparent veracity of my 1st-order reasoning to determine whether I took a drug that would make me unwittingly make mistakes. To do so would involve a question begging dismissal of this HOE.

But our Trick Questions Schema shows that we need not always bracket off our 1st-order reasoning when assessing the putative significance of HOE against that very reasoning. Rather, 1st-order reasoning can help to overturn the HOE provided by disagreement. Our basic contention is that, while 1st-order reasoning cannot overturn HOE against it by itself, it can be incorporated into an explanation that can. In TQS, 1st-order reasoning forms part of an explanation of why your peer made a mistake, but is nonetheless confident: she likely fell for the easy-to-miss trick.

We want to emphasize that 1st-order reasoning cannot overturn HOE by itself. Contrast instances of TQS with cases that patently beg the question:

---

5Another response would be to maintain Christensen’s bracketing criterion for HOE, but argue that peer disagreement provides another sort of evidence. You might think that because it’s harder to construct something like a TQS against Christensen’s drug example, our schema shows that there’s an important distinction between peer disagreement and these other sorts of examples.

6This may or may not be the only way in which 1st-order evidence can impugn HOE against it. We remain agnostic on this point, which is a question for future research.
Aha! Trick Questions, Independence, and the Epistemology of Disagreement

<table>
<thead>
<tr>
<th>Question Begging (QB)</th>
<th>Step One</th>
<th>In the course of your 1st-order reasoning, you first entertain a (non-tricky) premise, and then decide that it is incorrect.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Step Two</td>
<td>You have evidence that your friend used this premise in her reasoning, e.g. she got the answer you would get using this premise.</td>
</tr>
<tr>
<td></td>
<td>Step Three</td>
<td>You contend that, because she used the incorrect premise, your friend is mistaken.</td>
</tr>
</tbody>
</table>

This response begs the question because you merely assert that a step in your friend’s reasoning is incorrect because you think this step is wrong. You don’t explain why an epistemic peer would find the disputed step any more plausible than you do. In contrast, because the disputed step in TQS is a trick, it’s just the sort of mistake that might induce a false sense of confidence. And in particular, since you nearly fell for the trick yourself, you’ve seen first hand why someone of your epistemic caliber would fall for it. Consequently, unlike in QB, in Tricky Primes you’re not merely appealing to the veracity of your first-order reasoning to dismiss the evidence provided by disagreement. You avoid begging the question because you use your first-order reasoning to do something more: you incorporate it into an explanation of why, unlike yourself, your epistemic peer is confident in a bad piece of reasoning.

What’s crucial about the explanation available in Tricky Primes is that it explains a broadly psychological fact: namely, that your friend is persuaded by a false premise when you are not. Although we don’t have the space to develop a full positive proposal about what sorts of explanations can counter the HOE provided by disagreement, the above discussion suggests a constraint on any such proposal. To dismiss the evidence provided by disagreement without begging the question, you must provide an explanation that is *ad hominem*, in that it explains the psychological fact that your epistemic peer is persuaded by bad reasoning\(^7\). Of course, this is a necessary, but not sufficient condition for being justified in dismissing the evidence provided by disagreement (there are plenty of bad *ad hominem* arguments).

The upshot of our paper is this: Independence is false. In at least one respect, however, our departure from Christensen isn’t radical. We agree that you can’t always remain steadfast in the face of disagreement. But when you can remain steadfast, it’s because you countered the HOE provided by disagreement with the right sort of *ad hominem* explanation. And successful *ad hominem* explanations need not satisfy

\(^{7}\text{We think something like this principle underlies some common philosophical practices. When we disagree with talented opponents (such as brilliant philosophers of the past), we often feel rational pressure not only to explain which parts of their arguments are mistaken, but also why they were persuaded by that reasoning.}\)
Independence. Going forward, then, the task is to specify what makes this sort of *ad hominem* explanation successful. Our analysis of Tricky Primes suggests that, at the very least, an *ad hominem* explanation must explain your peer’s high credence in a bad piece of reasoning. Discovering what else is required of such explanations is a task for further research⁸.

**References**


⁸For their comments on earlier drafts of this paper, we are indebted to David Christensen, Gurpreet Rattan, Jonathan Weisberg, Andrew Sepielli, David Dyzenhaus, Oday Khaghani, Josh Brandt, Iain Laidley, and an anonymous reviewer from *Thought*. This research was supported by funding from the Social Sciences and Research Council of Canada.