PHI840: Intuitions and Conceptual Analysis

Week Eight – How to Beat a Counterexample

I want to talk for a bit about what moves we can make to defend a theory when faced with a putative counterexample. For simplicity, I'll be assuming we're defend the theory that all and only *F*s are *G*s. The following batch of moves seem, *prima facie*, to be available. (One interesting exercise is to see how many of these Lewis makes in "Causation as Influence", or even better, to discern other moves I have failed to notice.)

1. Deny that we have the Intuitions

This is perhaps the least plausible move to make in general, but if it can be made it is worth doing. I have already mentioned the example about the social aspects of knowing, which does seem to promote quite widespread disagreement. And certainly the agreement on the original Gettier cases is less than 100%. I'm not including here people like me who have the politically correct intuitions on Gettier cases but maintain that this doesn't *fatally* undermine reactionary standpoints.

Just to show this move can be worthwhile, I should mention the case where it seems most obviously applicable. There is some dispute about whether the sentence *Only x is F*, where *x* is a name, entails that *x* is *F*. So, more concretely, does the sentence *Only Mary voted for Carter* entail *Mary voted for Carter*? Clearly the first sentence communicates that Mary voted for Carter, but it is possible that this is because of the pragmatics of the sentence. (In fact it is because of the pragmatics of the sentence, but that's a topic for another day.) The linguist James McCawley argued that the inference must be pragmatic because it can be cancelled. (If you don't know what cancelling is, don't worry, much more on it below.) And the evidence that it could be cancelled is that the following sentence is felicitous[1]:

> Only Mary voted for Carter, and maybe even she didn't.

I have not found a single person who thinks McCawley is right about the pragmatics of this sentence. Everyone I have spoken to about it thinks the sentence is clearly defective. Now mistakes about where the intuitions lie are usually not as spectacular as this, but they can be made.

A more popular challenge than the flat-out assault is to question the depth of the intuitions. This is essentially what we do when we try to meet a counterexample by saying, "But isn't that case just like this one, and you agreed that my theory was right about this one, therefore, I'm right about that one too!" Jackson employs this in the opposite direction when he says that some cases we had thought were knowledge might be reconsidered in the light of the Gettier cases. And it's clearly one of the moves Unger makes in the attack on non-utilitarian intuitions. This is a useful methodology, but rather hard to employ, because it relies on some ingenuity in thinking up the nearby instances. So let's return to something easier.

---

[1] The argument is in his wonderful book *Everything Linguists have Ever Wanted to Know About Logic (But Were Afraid to Ask)*, University of Chicago Press, first ed 1981, second ed 1993. I think the argument is only in the second edition.

2. Detect a Change of Context

I suspect some people won't be familiar with Lewis's most recent response to the Kripke-Putnam argument for the rigidity of natural kind terms, so let's quickly run through it. This quote is from "Reduction of Mind", in the **Lewis** volume.

> Like any up-to-date philosopher of 1955, I think that 'water' is a cluster concept. Among the conditions in the cluster are: it is liquid, it is colourless, it is odorless, it supports life. But, *pace* the philosopher of 1955, there is more to the cluster than that. Another condition in the cluster is: it is a natural kind. Another condition is indexical: it is abundant hereabouts. Another is metalinguistic: many call it 'water'. Another is both metalinguistic and indexical: I have heard of it under the name 'water'. When we hear that XYZ on Twin Earth fits many of the conditions in the cluster but not all, we are in a state of semantic indecision about whether it deserves the name 'water'. When in a state of semantic indecision, we are often glad to go either way, and accommodate our own usage temporarily to the whims of our conversational partners. So if some philosopher, call him Schmutnam, invites us to join him in saying that the water on Twin Earth differs in chemical composition from the water here, we will happily follow his lead. And if another philosopher, Putnam, invites us to say that the stuff on Twin Earth is no water – and hence that Twoscar does not believe that water falls from the clouds – we will just as happily follow his lead. We should have followed Putnam's lead only for the duration of that conversation, then lapsed back into our accommodating state of indecision. But, sad to say, we thought that instead of playing along with a whim, we were settling a question once and for all. And so we came away lastingly misled (**Lewis**, 313-314).

Now Lewis admits that this move doesn't show the Putnam examples are worthless. For one thing, as Lewis admits, they show us things about the cluster that we didn't know before. For another, they show that conceptual claims we may have made about water are at least false on some acceptable disambiguations. But what Lewis's response makes clear is that some of these questions, like *Is water necessarily watery* may have no context-independent answer. This move can be employed to see off examples in theory of knowledge and conditionals, particularly when we note how easy it is to change contexts.

Apparently the following intuitions are widely enough held to warrant discussion. We know many commonplaces about the world. We know, for example, that we each have two hands. But we don't know that we are not brains in matrices, or holodeck images, or elements of whatever epistemic worst-case scenario you care to think about. But we also know that if we have two hands we are not in one of these worst-case scenarios. (Bracket for now wide-content concerns, and any details of the example which lead you to think we have hands in those cases.) So knowledge is not closed under known entailment. Lewis's response to this argument (at **Lewis**, 440) is to suggest that there is a change of context. Relative to an everyday context, it is right to say that we know we have two hands. Right to say, that is, because it is true. Relative to an epistemologist's context, it is right to say, and true, that we don't know we are in such a scenario. But in that context we also don't know we have two hands, because for all we know we are brains in vats, or whatever. And relative to the everyday context, we must know we are not brains in vats, because we know we have two hands, and hence we can rule out possibilities in which we are handless. By bringing up brains in matrices, or holodecks, we shift the context to one in which negligible possibilities are not neglected. In any context, knowledge is closed under known entailment, which is all that we originally wanted to prove.

A similar story works with conditionals, though here Lewis mistakenly endorses the other side. The following argument seems invalid, hence a principle we might call 'transitivity for conditionals' must be wrong. (As background, assume I am an avid skier, but not a suicidal one.)

If it snows this weekend, I'll go skiing.

If there's a blizzard this weekend, it will snow.

So, if there's a blizzard this weekend, I'll go skiing.

Intuitively we might have thought we could 'chain' conditionals together, so from If A, B and If B, C we derive If A, C. But, Lewis contends, this example shows that intuition was mistaken. A panoply of linguists, and a few brave philosophers, have pointed out that the proper response here seems to be to detect a change of context. (Indeed they were saying this long before Lewis said it about knowledge.) Relative to a context where blizzards are out of the question, it is true that if it snows, I'll go skiing. When blizzards are made salient, what is really true is that if it snows without being a blizzard, I'll go skiing. Remarkable evidence for this diagnosis ('remarkable' because there is almost never evidence for competing diagnoses in these cases) can be found by noting there are no intuitive counterexamples to transitivity where the premises are right-way-round, rather than back-to-front as in this example. This suggests it is the pragmatics which explains the data, not the semantics.

3. Inconsistent Intuitions

Clearly if the intuitions are inconsistent, they aren't all true. And since it is no cost to be at odds with intuitions which are false, showing this may escape a counterexample. This seems to be the strategy that Unger was pursuing in the passages we looked at.

As it stands, this move is a little quick. That the intuitions here are inconsistent just shows that one of them must be false, not that all of them are. (Indeed, it will sometimes show that one of them must be true.) What we must show is that the intuition that we want to drop is the false one in the set. Now sometimes this will be quite easy: if the folk are inclined to drop the intuition we want them to drop when presented with the inconsistency, this seems to be sufficient to seal the case.

Even still, this move is often misapplied. The criteria can't be that under any old presentation of the possible inconsistency the folk choose to resolve it by dropping the intuition which we want them to drop. Just as first-order intuitions can vary with the presentation of the material, so can second-order intuitions, or intuitions about which intuitions are stronger than others. Given the esoteric nature of the case, the possible influence of the interrogator on subject at this point is particularly large.

And this says nothing about the really hard case, the case where the folk are not sure what to do when presented with the inconsistency. (Compare what happens when the folk are presented with the semantic paradoxes.) I suspect the methodological moral is that this defence may fail unless it is quite clear that the intuition we want people to drop is the one which is least strongly held.

There may be another move which can be made to save the defence, though it isn't often appealed to, especially around here. It might be argued that the fact of the inconsistency shows that the folk are unreliable about these matters, and we all know that we shouldn't trust unreliable sources. The problem is that this casts doubt not only on the counterexample, but on our 'evidence' that All and only *F*s are *G*s, which will usually be little more than its strong intuitive plausibility. When we look at Horowitz's paper (in about 25 minutes) we might compare her strategies with this one.

4. Near Enough is Good Enough

The motivation for this move is the following passage from **Lewis**.

> Maybe nothing could perfectly deserve the name "sensation" unless it were infallibly introspectible; or the name "simultaneity" unless it were a frame-independent equivalence relation; or the name "value" unless it couldn't possibly fail to attract anyone who was well acquainted with it. If so, then there are no perfect deservers of these names to be had. But it would be silly to loose our Moorings and deny that there existed any such things as sensations, simultaneity and values. In each case, an imperfect candidate may deserve the name quite well enough (246).

The fact that the property picked out by *G* isn't a perfect deserver of the name "*F*" is no evidence that there is a distinction to be had between the *F*s and the *G*s; it might be that there is nothing which could be a perfect deserver of that name. This is very similar to the argument in my paper, so I won't spend much time on this point, other than to rehearse a quick summary of the arguments.

First, when there is no perfect deserver, an imperfect deserver will usually do. Second, there is reason to think that the kind of cases we discuss on philosophy are cases where there will be no perfect deservers. The reason is that if there were a perfect deserver, it would have been discovered long ago, and the issue would have ceased to be a live one. Finally, the kind of counterexamples we usually discuss in philosophy are, by the nature of the subject, liable to be the kind of extreme cases where imperfect deservers go imperfect. If *G* imperfectly deserves the name "*F*", there will have to be some cases where intuition says that *a* is *F* and *a* is *G* differ in truth value. That's just what it is to be imperfect. It is better, *ceteris paribus*, for these cases to be extreme cases which are hard to think up and on which intuitions are not always unified or clear. In other words, once you've accepted that your analysis is imperfect, the last thing you should be worried about is a philosopher's counterexample.

5. Scalar and Absolute Predicates

This is a move Lewis makes at a couple of crucial points in the causation paper. On Lewis's theory the presence of any object in the near-ish vicinity is a cause of every event, because of the gravitational and electro-magnetic forces objects bring with them. This is very odd, since it makes almost every object a cause of almost every event, since those forces tend to dissipate over rather large distances. Lewis's response is to accept that this is, in a sense true: the spatio-temporal location of any particular object is a cause of any particular event. But he denies that it is much of a cause. What we thought was an absolute, or on/off predicate, turns out to be essentially a scalar predicate. These weak attractive forces are very small causes of events; something is properly called a cause if it a sufficiently large cause. See page 15 of the paper for a discussion of this point, replete with some questionable analogies to similar moves regarding quantifiers.

The general point is that if there is a confusion between scalar and absolute uses of a predicate, the folk may well confuse something not being much of an *F* for it not being an *F*.

6. False Implicit Theory

Jackson says at a couple of points that we can give less weight to an intuition about a possible case if we can show that the 'intuition' isn't basic, but is derived from some more general intuition. This seems clearly wrong in general; why not say general intuitions are more likely to be right? But if we can show that the general intuition is wrong, and

that the only reason for the strong view about the possible case this view is forced by the general intuition, we seem to have won the game.

This may be the case in debates about personal identity across time and worlds. If someone has the intuition that the property 'being part of the same person' is intrinsic to a pair of stages, they will have all sorts of odd intuitions about hard cases in the identity literature. Since the intrinsicness intuition is (a) plausible and (b) false, recognising this could lead to placing a more appropriate weight on the folk intuitions.

This is less of a risk, but I guess many folk give some credence to the idea that it is impossible to have beliefs and desires without having conscious states. I don't know how plausible this is; it certainly sounds false to me. Anyway, perhaps this theory is behind the anti-functionalist intuitions in Chinese room and nation cases in philosophy of mind.

7. Mistaken Identity

This is to some extent a variant on the previous defence. Sometimes it is possible to show that an implicit theory is plausible but false by showing that there is a true theory with which it is easily confused. For instance, utilitarians deny that it is always wrong to execute innocents. But they agree that it is always wrong *ceteris paribus* to execute innocents, and that it is almost always wrong in practice to execute innocents, and so on. This can become important in metaphysics when there is the possibility for subtle confusions between intuitions of metaphysical possibility and intuitions of epistemic possibility.

For example, it seems possible that I could retain my identity while losing all the properties that are normally taken to be constitutive of my identity over time. There doesn't seem to be a contradiction in the story that runs: "Brian woke up one morning with an entirely new body and no memories of his former life. Had he been able to remember that he was on the run from the Mafia, he would have been quite pleased with this evasive technique." If personal identity over time goes by psychological continuity, or physical continuity, or some combination of the two, this story is incoherent. So why don't we take this clear possibility to refute those theories of identity. One possible out is to say that the reason this story seems possible *simpliciter* is that a Cartesian theory of identity, that identity means preservation of soul-stuff, is epistemically possible, and if that theory is true then the story I told is metaphysically possible.

As a general point, and this really can't be stressed enough, the more you can say to explain why we have intuitions which are, on your theory, false[2], the better off your theory is. It's no solution to just say, we have these false intuitions and that's the end of it. The cheapest, and these days least interesting, way to do this is by appeal to evolutionary history. Showing that we're trying to grasp something which just ain't there, like a complete arithmetical theory with recursive functions, or an epistemic standard between justification and certainty, is also a way out. (This is one description, possibly a good description, of what we do when we show intuitions are inconsistent.) What we do here is claim that our reports on what the intuitions really are is theory-laden and hence possibly wrong. The intuition is really that something like *blah* is true, and my theory satisfies this. So Putnam's essentialism about water is squared with the intuition that the stuff which fills the oceans *etc* might not have been H_2O by distinguishing between two senses of 'might', noting that his essentialism is only incompatible with one sense, and the satisfaction of the other sense is good enough to satisfy the intuition. As a different example, this is

[2] Some may have the intuition that intuitions don't have propositional content; that would be another false intuition.

why the utilitarian talks about how Kantian morality usually maximises utility in the actual world. This issue also comes up a lot in connection with vagueness, but that's for (at most) a later seminar.

8. Guilt by Association

I mentioned earlier that if we can show intuitions are unreliable in a particular area, that is a reason not to trust them.[3] There are tricky questions, which we have been looking at, about how we can show this, and indeed about just what it would be to show this, but we will presume for our purposes that those have been settled. This is (at least part of) what Horowitz in trying to do in her paper in **Intuition**. What we'll be interested in are two questions. First, when can this kind of argument be used against a counterexample? Second, is this one of those cases?

To try and get a feel for this kind of case, and for a quick introduction to some of the probabilistic issues facing us, I want to first look at a slightly similar argument. One of the most central principles of modern decision theory is what I'll call here Dom, for Dominance Principle. We have an intuitive idea of the value of a bet. (In decision theory every action is reduced to a bet.) And we also have a not too bad idea of the conditional value of a bet, how much a bet would be worth were it to be the case that some condition is met. In simple cases this is easy to work out. This betting ticket worth \$10 is worth \$2000 if Hopeless Horse wins and \$0 otherwise. In other cases it is a little harder. A bet to win \$100 if John McCain is the next President is worth virtually nothing if he loses the Republican Primary, but still worth less than \$100 if he wins that primary. Formally, let $V(A)$, where A is an action, be the value of A, and $V(A \mid p)$ be the conditional value of A if p is true. In symbols, Dom says the following inference is valid.

$V(A \mid p) > V(B \mid p)$
$V(A \mid q) > V(B \mid q)$
$Pr(p \veebar q) = 1$

$V(A) > V(B)$

Note that $\veebar$ stands for exclusive disjunction. So if you're better off choosing A should p happen, and better off choosing B should q happen, and it is certain that exactly one of these will happen, then you are better off choosing A than B. This principle is obviously important for the two-box argument in Newcomb's Problem, as we discussed a few weeks back. Now Dom is *very* intuitive; I think that there is no argument for it except its intuitive plausibility. (There are a few other purported arguments for it in the literature, but really I think these are little better than intuition pumps. As arguments they rely on premises less plausible, and certainly more controvertible, than Dom itself.) But the intuitions which support Dom also support two rather dubious principles, Dom-Inc (for Inclusive) and Dom-Inf. Each of these seem susceptible to knock-down counterexamples. In symbols, here are the principles.

Dom-Inc

$V(A \mid p) > V(B \mid p)$
$V(A \mid q) > V(B \mid q)$
$Pr(p \vee q) = 1$

$V(A) > V(B)$

[3] This obviously is not an endorsement of reliabilism!

Dom-Inf

$$\frac{V(A \mid p_1) > V(B \mid p_1)\ \&\ \ldots\ \&\ V(A \mid p_n) > V(B \mid p_n)\ \&\ \ldots \quad Pr(p_1 \veebar \ldots \veebar p_n \veebar \ldots) = 1}{V(A) > V(B)}$$

Here's the problem case for Dom-Inc. I'm going to deal you two cards from a three card deck. The three cards are the ace of spades, the ace of clubs and the two of hearts. Assume I have now dealt them and the cards are face down in front of you. The remaining card is face down in front of me, and I haven't been able to see it. I then offer to sell you a bet which will pay $10 if you have both aces for a bargain price of $4. Is this a good buy?

If you are hesitant, I try and talk you into buying with this argument. If you have the ace of clubs, there is a 50/50 chance that you also have the ace of hearts. So if you have the ace of clubs, this bet should be worth $5. And if you have the ace of hearts, there is a 50/50 chance that you also have the ace of clubs. So if you have the ace of hearts, this bet should be worth $5. But it is certain that you either have the ace of clubs or the ace of hearts, so by Dom-Inc, the bet is worth $5. Are you convinced?

Here's the problem case for Dom-Inf. A demon plays the following game. He starts with a fair coin and two envelopes and tosses the coin a number of times until he gets heads the first time. If it lands heads he puts $3 in the first envelope. If it lands heads the second time he puts $9 in the first envelope, the third time $27, the fourth time $81 and so on. If it never lands heads after an infinite number of throws he puts $3 in the first envelope. (Demons can perform infinitely many acts, and have large bank accounts.) He then tosses the coin again and if it lands heads he puts 1/3 as much in the second envelope as in the first, and if it lands tails he puts 3 times as much in the second envelope as the first. He then gives you one of the envelopes, though he won't say which. But that's the last envelope you'll get for free. He now wants you to sell you the chance to swap envelopes for $1. We won't discuss the merits of that trade, but let's consider the question of whether your envelope, which you should now sign before the demon starts working his spells, is worth more or less than the other envelope. As usual, it depends on whether you are a pessimist or an optimist. As less than usual, whether you are a pessimist or an optimist depends on how you partition reality.

I won't bore you with the math, so you'll have to trust me that the following is true. Let p_1 be the proposition that your envelope has $1 in it, p_2 that it has $3 in it, p_3 that it has $9 in it, and so on. And let B be your envelope and A the other envelope. The following are all true: $V(A \mid p_1) > V(B \mid p_1)$, $V(A \mid p_2) > V(B \mid p_2)$, $V(A \mid p_3) > V(B \mid p_3)$ and so on. Hence by Dom-Inf, $V(A) > V(B)$. As usual the world is a mean place. But wait! Let q_1 be the proposition that the other envelope has $1 in it, q_2 that it has $3 in it, q_3 that it has $9 in it, and so on. Then the following are all true: $V(B \mid q_1) > V(A \mid q_1)$, $V(B \mid q_2) > V(A \mid q_2)$, $V(B \mid q_3) > V(A \mid q_3)$ and so on. Hence by Dom-Inf $V(B) > V(A)$. And that looks like a contradiction to me.

So by parity of reasoning, Dom is false. Well, it had better not be, because without it we know **ABSOLUTELY NOTHING** about decision theory. I suppose knowing absolutely nothing about an area isn't considered too bad in philosophy. (How much do we really know about metaphysics after all these centuries?) But in a real science, like decision theory(!), knowing absolutely nothing is a **VERY BAD THING INDEED**. It does seem very odd to deny instances of Dom. But not, to my ear, any odder than it sounds to deny instances of Dom-Inc or Dom-Inf, and they lead to mistaken conclusions. So intuitions around here are unreliable, so we shouldn't trust our intuition that Dom is correct. Maybe I'm just weak-willed to go on accepting it, knowing as I do that I have no evidence for it.

9. Horowitz's Argument

- Start by looking at the Kahneman and Tversky experiments.

  Why are the actions irrational?

  Why do we draw these implications out of them? (Are we still maximising something)

- Compare these examples to the Doing and Allowing examples

  Is there a close analogy?

  If so, what are the implications of this?

- What happens if we alter each example a small amount?

  If this makes the cases diverge, does this show the original cases were different.

- Why does this explanation of the epistemic cases affect what we say about moral cases?

  It can't be that if we can explain moral judgements they are worthless

  So it must be that there is something particular to this kind of explanation.

10. Grice on the role of pragmatics in metaphysics

- What are our intuitions about counterexamples intuitions of?

  Perhaps of defectiveness

  That is, our first reaction is: I wouldn't say that

  But we don't say things for lots of reasons: here's three

  Ungrammatical

  False (semantically defective)

  Pragmatically false (misleading)

  Only the middle one matters to metaphysics – who cares if 'He has good handwriting' is misleading

- How can we tell which of the three?

- Generating pragmatic implications

  Can't just say, "Oh this is part of the pragmatics"

  So Grice gives us a theory, in the first few pages of chapter 2

- How breaches play out

  Some breaches are accidental, b/c we make mistakes about what is relevant

  Some breaches are deliberate, lying

  Some breaches are explicit, saying that we are not playing the game

  And sometimes we 'flout' the maxims

- How this matters for metaphysics

  I perceive *M* means *M* causes my having certain sense-data in the right kind of way

  To be having a red sense-data is to be such that something looks red to me.

  But when I'm standing in front of a British post-box, I don't have the 'looks red' sensation

  It would be horribly misleading to say this 'looks red'.

  Hence it doesn't look red, but I am perceiving a red thing, so the CTP is wrong.

  Mistake: inferring from *S* is horribly misleading to *S* is false.

- How can this technique be used?
- Consistency proofs and correctness

  All the Grice story shows is that we don't have a counterexample here

  But there are many other theories which are also immune to counterexamples

  How do we tell between them

  This is a big question: I hope simplicity will be the answer

- Objections to the Gricean picture

11. The Gricean Maxims

Quantity	1. Make your contribution as informative as possible.
	2. Do not make your contribution more informative than is required.
Quality	1. Do not say what you believe to be false.
	2. Do not say that for which you lack adequate evidence.
Relevance	1. Be relevant
Manner	1. Avoid obscurity of expression.
	2. Avoid ambiguity.
	3. Be brief (avoid unnecessary prolixity)
	4. Be orderly.

12. Data Sentences

(1) Everyone has heard of the counterfactual analysis of causation.

(2) I don't like cricket, I love it.

(3) Alice: Did any of the passengers die?

Bob: Yes / *Some did

(4) Steffen comes from Norway, or somewhere in Scandinavia.

(5) *Steffen comes from somewhere in Scandinavia, or from Norway

(6) If Brian has several drinks and drives home, he's a very irresponsible driver

(7) If Brian has several drinks and drives home, he's a very anti-social drinker

13. How Gricean Implicatures are Generated

P1 S uttered a sentence with a particular meaning, in a given context

P2 S is observing the co-operative principle

P3 Given P1, S could not be observing the cooperative principle unless he believed *p*

C S implicated *p*

This is from Wayne Davis, *Implicature*, Cambridge University Press, page 15.

14. For next week

We will be returning to the Jackson book, looking at the rest of chapter two and of chapter three. Despite what is listed for week nine, the official reading really is Jackson pages 44 to 86, and the Stalnaker paper "Assertion" which is **already** in the filing cabinet!