Handout 12: The Hawthorne Problem, Backwards Counterfactuals . . . And Some Metaphysics!

Philosophy 691: Conditionals
Northern Illinois University ⋆ Fall 2011
Geoff Pynn

THE HAWTHORNE PROBLEM

1. Consider an instance of $A \boxrightarrow C$ that seems clearly true. (a) and (b) imply that $A \boxrightarrow C$ is false:

   (a) $A \boxrightarrow C$ is true iff at all the closest A-worlds, C.

   (b) There is an $(A \,\&\, \neg C)$-world among the closest.

   The example Hawthorne uses for A and C are 'I dropped the plate' and 'The plate falls to the floor'. The motivation for (b) is that according to one interpretation of quantum mechanics, there was a (very very tiny) chance that the dropped plate would fly off sideways. But we could use a non-quantum-mechanical example instead, at least when initially stating the problem. DeRose's case: A = 'I tagged up' and C = 'We won the game'; then (b) is motivated because it seems that there was a (much less tiny, in this case) chance that when tagging up I would trip and fall.

2. The first solution is to replace 'all' with 'most' in (a). This solution invalidates Agglomeration:

   > AGGLOMERATION. If $A \boxrightarrow B$ and $A \boxrightarrow C$, then $A \boxrightarrow (B \,\&\, C)$.

   The second solution is to say that the relevant $(A \,\&\, \neg C)$-world is not among the closest. Most of Hawthorne's paper is devoted to showing how the criterion that Lewis suggests for doing so (roughly: if w_1 contains a 'quasi-miracle' that w_2 does not, then (ceteris paribus) w_1 is further from α than w_2) leads to bad results.

3. I want to grant that the quasi-miraculousness criterion (as suggested by Lewis and fleshed out by Hawthorne) won't work. And I would like to respect Agglomeration. So, what should we do?

4. Hawthorne suggests the following solution:

   > [M]y own preference is to opt for a picture according to which, for any possibility that P, and any world w, there is a unique closest world to w where P. [...] From this perspective, matters are obviously very different when it comes to the problems at hand. Suppose I do not drop a plate at t and the world is chancy. There is a closest world where I drop the plate at t. If it goes off sideways at the world the counterfactual is false. Otherwise it is true.

   One way of stating this picture is in Stalnakerian terms: $A \boxrightarrow C$ is true at α iff at $s\langle\alpha, A\rangle$, C, but (now departing from Stalnaker) we say that s does not take different values in different contexts, but that there is a single 'correct' value for $s\langle\alpha, A\rangle$.

5. In addition to Hawthorne's solution, here are three other ways you might approach the problem:

   (a) *Skeptical.* $A \boxrightarrow C$ is false. But there is a consolation prize. Let Ch be a probability function; $Ch(P)$ states the objective chance that P. And $A \boxrightarrow (Ch(C)$ is extremely high$)$ is true. Plusses: we do not need to deny (a), (b), or Agglomeration. Minuses: many (how many? I'd think very, very many) of the counterfactuals we ordinarily treat as true are false. Does the consolation prize mitigate this minus?

(b) *Loose Talk.* Suppose that strictly speaking (a) is correct; given (b), $A \boxrightarrow C$ is strictly speaking false. But we often speak loosely, not strictly. There are two ways to develop this idea:

First, we might say that $A \boxrightarrow C$ is true only if it is strictly speaking true. But in loose contexts, it's okay to falsely assert $A \boxrightarrow C$, because by falsely asserting that $A \boxrightarrow C$, we communicate (something like) that $A \boxrightarrow (Ch(C)$ is extremely high). This way of thinking about the loose talk approach is a way of developing the skeptical resolution above.

Second, we might say that what an utterance of '$A \boxrightarrow C$' expresses, and hence whether or not the utterance expresses something true, can depend upon whether we are speaking strictly or loosely. In loose contexts, someone who utters '$A \boxrightarrow C$' may assert something true, even though in strict contexts the same utterance would constitute a false assertion. This is a form of contextualism about '$A \boxrightarrow C$': it says that the sentence form '$A \boxrightarrow C$' can be used to assert truth-conditionally distinct contents in different contexts, even holding A and C fixed. One challenge is to explain the parameters that determine the truth-conditional content asserted by an utterance of '$A \boxrightarrow C$' relative to a context. Some candidates for context-sensitivity:

i. The closeness relation?
ii. The domain restriction on the quantifier 'all' in (a)?
iii. Replace 'all' with 'most' in (a); what shifts is the threshold for 'most'?

(c) *Pragmatic Agglomeration.* Replacing 'all' with 'most' in (a) invalidates Agglomeration. But perhaps a pragmatic rule can explain why Agglomeration seems valid. 'Most' is vague. In general, if F is a vague predicate, you should not assert that Fa when a is too close to being an indeterminate case of an F. Applied here, the general principle gives us the rule:

> Assert $A \boxrightarrow C$ only if it is not too close to being indeterminate whether most A-worlds are C-worlds.

How close is too close? That's a good question (with an undoubtedly vague answer!), but there is reason to think that the answer will imply that it is okay to assert $A \boxrightarrow B$ and $A \boxrightarrow C$ only if $A \boxrightarrow (B \,\&\, C)$ is true. Suppose that the vague threshold for 'most' is centered around 95%. Even if it is determinately true that 96% is most, 96% is too close to being indeterminate; the rule says don't assert $A \boxrightarrow C$ when merely 96% of closest A-worlds are C-worlds (even if the assertion would be true). Note, too, that if 96% of closest A-worlds are B-worlds and 96% of closest A-worlds are C-worlds, then it might be that only 92% of closest A-worlds are $(B \,\&\, C)$-worlds; this would be a case where Agglomeration fails. On the other hand, 99% seems like a safe enough case of 'most'. And if 99% of closest A-worlds are B-worlds and 99% of A-worlds are C-worlds, then no fewer than 98% of closest A-worlds are $(B \,\&\, C)$-worlds.

BACKWARD COUNTERFACTUALS

1. Most of our $A \boxrightarrow C$ thoughts are 'forward'; i.e., the time relevant to the truth of the antecedent is earlier than that relevant to the truth of the consequent. But not all; e.g.:

   B If she had arrived in Chicago at five, then her train would have left Elburn at 3:30.

   That seems like a fine thing to say and believe, but it is 'backward'. Some people say that 'backward' counterfactuals require 'would have to have' instead of 'would have'. I don't think B needs that modification to be felicitous, but it certainly *could* take it. This is an interesting phenomenon. Bennett suggests that the 'would have to have' indicates that the consequent is (part of?) the *best explanation* of the antecedent. I have trouble understanding this, but that is probably due to my idiosyncratic difficulties with the concept of explanation. I'm happy to concede, though, that the 'have to have' sounds like it's adding something epistemic to the consequent, similar to some uses of 'must'; e.g. when I assert 'It must be raining' upon seeing people enter the building wearing wet raincoats.

2. Bennett's ramp-n-fork theory of closeness, developed to account for ordinary forward counterfactuals, appears to do a nice job of giving truth-conditions for backwards counterfactuals too. Recall that for Bennett,

w is a closest A-world to α just in case (a) at some time t_F shortly before t_A, w forks slightly from α, and (b) w evolves smoothly from t_F through t_A, making A true, and then beyond. Suppose $A \boxrightarrow C$ is backwards; then Bennett's thought is that $A \boxrightarrow C$ is true iff all of the forks on the closest A-worlds run through C. That seems right for B above.

3. Little question: given ramp-n-fork, it seems like backward counterfactuals where C is an actual truth in the distant past should all come out true; after all, if all the closest A-worlds resemble α perfectly before t_F, then any pre-t_F truth at α will be true at all those worlds, too. 'If I had eaten Wheaties this morning, then the Battle of Hastings would [still] have been fought in 1066.' Does that seem true to you?

4. The Bennett account predicts that $A \boxrightarrow C$ does not imply $C \boxrightarrow A$, which is how it should be. His example is good: 'If my father had asked me in his will to walk a mountain trail in honor of his 100th birthday, I would have done that' can be true even if 'If I had I walked a mountain trail in honor of my father's 100th birthday, then his will would have asked me to do that' isn't. All the closest forks that lead to Will also lead to Walk, though none of the closest forks that lead to Walk reach back as far as the time relevant to Will.

COUNTERPARTS

1. In this puzzling but intriguing little section, Bennett suggests several interesting ideas. Here I will attempt to reconstruct one of them.

2. We are all convinced, pre-theoretically, that particular things could have differed in certain ways from how they actually are; e.g., that I could have been skinnier today than I actually am. Translated into possible worlds talk, this obvious truth becomes:

   S There is a possible world where GP is skinnier today than he actually is.

   Lewis, mad genius, held that if 'GP' is understood as referring to the actually existing person Geoff Pynn, then S is false, since on his view GP is 'world-bound'; i.e., exists at the actual world and no other world. He was driven to this bizarre idea by his extreme realism about possible worlds together with some not so crazy ideas about what individuals like GP are. Other deranged ideas would also lead to the same place; for example, Leibniz's view that everything that can be truly predicated of an individual is essential to it.

3. (Quibble: on page 281, Bennett denies that this was Leibniz's view. He says that Leibniz thought merely that every intrinsic property of a thing is essential to it. This is deranged enough to force Leibniz to deny S, provided skinniness is an intrinsic property, which it seems to be. But I invite you to consider the following passage from Leibniz's *Discourse on Metaphysics* and see whether you think it squares with Bennett's interpretation:

   > [W]e can say that the nature of an individual substance or of a complete being is to have a notion so complete that it is sufficient to contain and allow us to deduce from it all the predicates of the subject to which this notion is attributed. [...] God, seeing Alexander's individual notion or haecceity, sees in it at the same time the basis and reason for all the predicates which can be said truly of him, for example, that he vanquished Darius and Porus; he even know *a priori* (and not by experience) whether he died a natural death or whether he was poisoned, something we can know only through history. Thus when we consider carefully the connection of things, we can say that from all time in Alexander's soul there are vestiges of everything that has happened to him and marks of everything that will happen to him and even traces of everything that happens in the universe, even though God alone could recognize them all (*Discourse* 8).

   You can read on:

   > [E]very substance is like a complete world and like a mirror of God or of the whole universe, which each one expresses in its own way... [Every substance] expresses, however confusedly, everything that happens in the universe, whether past, present, or future (*Discourse* 9).

It seems to me that Leibniz is here saying that the nature (i.e., the essence) of an individual substance grounds *all* of its properties and relations, and not just its intrinsic properties. This *may* be compatible with denying that its extrinsic properties are essential to it (depends on what Leibniz means by 'contain', 'there are vestiges of and marks of', 'expresses', and what I mean by 'grounds'). But the much more natural reading is that Leibniz is saying just what Bennett says he doesn't. End quibble.)

4. Though he denied S's literal truth Lewis held that a closely related proposition could be true:

   S_L There is a possible world where GP's counterpart is skinnier today than GP.

   At a first pass, X's counterpart in w is whatever plays the X-role in w. If you think the question, 'Which possible worlds are closest?' is tough, wait until you think about the question, 'Which possible worlds contain counterparts of X?'

5. But those who wish to say that S is literally true (and not merely true-lite in virtue of S_L's truth) may wind up at a lonesome destination. Here is the road. In virtue of what is S true? What makes it the case that the actual GP could have existed while lacking a property he actually has? The question is made more poignant by noting that we do not wish to affirm that GP could have had just *any* property he doesn't actually have. Though S is true, I do not wish to affirm that, and am even inclined to deny that:

   O There is a possible world where GP has different biological parents than he actually does.

   What explains the disparity between S and O? The lonesome destination is reached when you find yourself thinking: S is true and O false because of GP's Essence or Nature or Kind—some Deep Metaphysical Fact about GP that explains which properties GP could have had and which he could not.

6. Some philosophers, their cognition dulled by prolonged exposure to neo-Scholastic metaphysics, will not blanch at Essences. But others among us, upon hearing them afresh, will be struck by how primitive and mysterious they sound. Can we affirm S, refrain from affirming O, and avoid taking this road? Faced with the choice between counterparts and Essences, the former might not look so bad. The counterpart theorist can explain the disparity in our attitudes between S and O: while it is easy to conceive of a possible world where a skinnier person plays the GP role, in most possible worlds we can conceive where GP's parents do not have GP, there is no one to play the GP role.

7. Bennett offers a basis for affirming S without affirming O that purports to steer through the Scylla of counterparts and Charybdis of Essences. Suppose that $A \mathrel{\Box\!\!\rightarrow} C$ is true, and let t_E be the time of the earliest fork among the closest A-worlds. Before t_E, each of those worlds resembles the actual world perfectly, so anything true of X at the actual world pre-t_E was true of X at each of those possible worlds pre-t_F. If X actually exists pre-t_E, then X exists at all the closest A-worlds. Thus the existence of X (at some time) is compatible with the truth of A. My basis for affirming S is that plenty of counterfactuals whose antecedents imply that GP is skinner than he actually is are true.

   But if at (some of) the closest A-worlds, X does not exist prior to t_E, then we have no secure way of saying whether X's existence is compatible with the truth of A. 'If my actual biological parents had never had children...' takes us only to worlds which fork from α before I exist, and so we have no secure way of saying whether I exist at those worlds. Hence our discomfort with O.

8. So Bennett's idea is that we can explain the truth of claims like S without reference either to counterparts or to Essences. The explanation predicts (correctly) that we should have no inclination to affirm O. Here is a test for the explanation: take a claim of the form 'X could have been F'. Find a true counterfactual whose antecedent would imply that X is not F but where the earliest forks at the closest A-worlds are after X has come into existence. Then you should think that the claim is true. If you can't find such a counterfactual, you shouldn't.

9. Suppose you can't. You might still want to know: okay, but *could* X have been F? Bennett counsels us that such questions are "empty, pointless: no answer has a respectable basis, nor do our serious intellectual needs and interests require us to find one" (p. 283).