

Probabilities; free will; history versus laws

October 24 // Seminar on Ability // NYU, Fall 2024

Matt Mandelkern

1 Probabilities and puzzles for the ACA?

Recall:

- (1) Ann is handed two fair decks of cards. What is the chance that she can draw a clubs from one of the decks without looking?

Building on similar cases suggested to me by Ben Holguín and an anonymous reviewer.

Intuitively, there are two judgments available:

- $\frac{1}{4}$ (the chance that she draws clubs, conditional on trying to);
- 1 (the chance that she will draw a clubs, conditional on trying to draw c , where c is any clubs card in the deck).

The ACA predicts both judgments, depending how the context chunks up the practically available actions. But the ACA also predicts other judgments, e.g. when the context divides up the available actions as:

{draw a card from Deck 1, draw a card from Deck 2, don't draw a card from either deck}

it predicts that the chance that Ann will be able to draw a clubs is $\frac{7}{16}$, which does not seem to be available. But compare (2):

- (2) One of the decks is such that, if Ann tries to draw a card from it, she'll draw a clubs from it.

Many people say this has chance $\frac{1}{4}$. So the puzzle may be about intuitions about probabilities of quantified conditionals, not the ACA.

This kind of case is due to an anonymous reviewer.

- (3) *[There are ten buttons, numbered one through ten, exactly one of which (say, seven) will activate auto-pilot. Jim doesn't know which button turns on auto-pilot.]*
- a. What is the chance that Jim can now engage auto-pilot?

The ACA can predict $.1$ and 1 . But it apparently predicts other readings, too, distinguishable by their probabilities:

- Suppose the available actions are *{press one or two, press three or four, press five or six, press seven or eight, press nine or ten, don't press a button}*. The chance that one of *these* actions is such that, if Jim tries to do it, he'll engage the auto-pilot, is plausibly $.5$.
- or if we have *{press an odd button, press an even button, don't press a button}* we get a chance judgment of $.2$.

Etc. Compare again the overt quantified conditional:

- (4) There is an action such that if Jim tries to do it, he'll engage the autopilot.

The salient readings of (4) seem to be ones where the chance is .1 or 1, not the intermediate ones. If so that suggests the fault is not the ACA but the way we quantify over actions: we do so in either a maximally fine-grained or maximally coarse-grained way.

Of course, this is very vague.

2 Non-agents, again

Probability judgments may help resolve a puzzle about cases where we apparently ascribe abilities to non-agents:

- (5) This elevator is able to carry three thousand pounds.
 (6) This black hole is able to absorb that galaxy.

from Irene Heim, attributed to Maria Bittner

This is an objection to any form of conditional analysis, since e.g. (6) doesn't mean that the black hole will absorb the galaxy *if it tries*. The deeper question is whether we have any truly *agential* modality at all, or everything is in some sense about (im)possibility.

Probability judgments suggest that these cases are actually different: (5) is an ability ascription, where the trying is done by a covert, generic agent, while (6) is a circumstantial modal.

Suppose that conditional on loading the elevator with three thousand pounds of cargo, there is a 30% chance that the cord will snap, and a 70% chance that the elevator will work as normal. In that case, the probability of (5) is intuitively, 70%. That is, credences again seem to track conditional probabilities. That suggests an analysis of sentences like (5) along the lines of a conditional analysis, but with a covert generic agent.

By contrast, in the case of (6), appealing to a covert generic agent obviously won't help. But the case also seems totally different from all the cases of ability ascriptions we've looked at so, whose probabilities always matched a salient *conditional* probability judgment. But this doesn't seem to be true in this case.

What *should* your credence in (6) be? It seems like it should just track your credence that there is *some* possibility that the black hole absorbs the galaxy.

3 Freedom and the CA

Let us say that determinism is true at our world. Familiar arguments purport to show that, if this is the case, then no one has the ability to do

anything, except perhaps for what she actually does. . . . But if (CA) is true, then agents would have the ability to perform various actions that they do not actually perform. For it is plausible that the conditionals in terms of which (CA) analyzes ability would still be true in a deterministic world. But then, since it makes false predictions about such a world, which for all we know may be our own, (CA) is false. SEP, *Ability*

I sort of agree that if the CA is true then compatibilism is true, where:

Soft determinism is the doctrine that sometimes one freely does what one is predetermined to do; and that in such a case one is able to act otherwise though past history and the laws of nature determine that one will not act otherwise. Compatibilism is the doctrine that soft determinism may be true.

Assume determinism; could one have done otherwise? Well that just concerns the truth of a certain conditional — but who ever thought that determinism implies that all counterfactuals are false?

(The details of the conditional analysis don't matter for this point.)

I think it's dialectically bizarre to go from the inconsistency of a CA with incompatibilism to a rejection of a CA.

I'm not sure I understand the exact role of the CA here since on *any* analysis ability ascriptions concern *how things could have gone otherwise*. If determinism is true, then, *holding fixed how things actually were*, things couldn't have been different. But that's also true if determinism is false.

Lewis, *Are we free to break the laws?*

Set aside issues about *freedom* which seem heady and obscure.

3.1 Changing history or the laws?

Suppose there is some initial conditions H and laws L such that $H \& L$ entails φ . Could it still be true that $A_S \neg \varphi$? Given the CA: could it still be true that, if S had tried to $\neg \varphi$, she would have succeeded?

Surely there is no general reason to think not, any more than there is a general reason to think that when $H \& L$ entails φ , $\neg \varphi \supset \psi$ could never be true. Conditionals take us to *different possible worlds*, where some actual history or law is different.

But there's a residual puzzle. Supposing determinism is true and φ happens, then any world where $\neg \varphi$ is true is one where $H \& L$ is false. Does it follow that you could have (broken a law or changed the past)?

- we could have violated an (actual) law.
- the initial conditions / distant past would have been different.

Dorr, *Against counterfactual miracles*.

Either way, we can argue that you didn't *cause* the change in question.

And in both cases, it's not like you broke the laws or changed the history *in the world where you act*. Right now, I'm breaking laws of nature that are true elsewhere. No one treats me like some kind of god.

Lewis worries about counterfactual dependence, but maybe that's not true.

4 Laws or history?

There's still an interesting general question: what would have been different, had something gone differently? In a deterministic universe where $H \& L$ is true, suppose you don't lift your hand.

(7) If I had lifted my hand, $H \& L$ would have been false.

This seems undeniable; $H \& L$ entails the negation of the antecedent. Of course, on most theories of the conditional, $p > \neg p$ can be true, when p is impossible. It's a little hard to see someone defending (7), though.

In Stalnaker's logic, (7) entails:

Note this doesn't follow in Lewis's logic.

(8) If I had lifted my hand, $\neg H$; or, if I had lifted my hand, $\neg L$.

Which is it?

4.1 Laws

Lewis (1973) wanted to give an account of counterfactuals in terms of (non-modal) "similarity", which seems to face prima facie obstacles from observations like Fine's:

(9) If Nixon had launched nuclear weapons, there would have been a nuclear holocaust.

The world most like actuality where Nixon launches nuclear weapons is plausibly one where they fail to detonate somehow.

Lewis (1979) tried to fix things up with a story on which history and laws would have remained the same until shortly before the time of the counterfactual antecedent. Then a "small miracle" would have occurred, violating laws (and changing the *recent* past).

Intuition: the future depends counterfactually on the past, but not v.v. This is the sense in which the future is open but the past isn't.

(10) If I were home right now, I would be shocked and confused.

This seems false, but if a small miracle had moved me from my classroom to my apartment, I would be shocked and confused. So we need an 'on-ramp' to the miracle.

(11) If I had lost my coat last year, I would have lost it on Dec. 31.

It seems there's no reason to think (11) is true.

And what about 'back-tracking' cases?

(12) Jack and Jim are in a fight. If they had both come to the party, however, they wouldn't have fought, since they would only

both have come if they hadn't gotten into a fight in the first place.

Lewis wants to set aside these cases as "non-standard", and that's been a common move, though I'm not sure what justifies it.

Lewis motivates his view as part of a bigger research program which aims to reduce causation to counterfactuals and reduce counterfactuals to non-modal facts.

Worlds [with small changes throughout history] should not turn out to be the most similar worlds. . . That would lead to back-tracking unlimited. (And as Bennett observes. . . it would make counterfactuals useless; we know far too little to figure out which of them are true under a resolution of vagueness that validates very much back-tracking.) . . . a lot of perfect match of particular fact is worth a little miracle.

4.2 History

Dorr (2016) sets up the puzzle this way:

Normal here: lacking godlike powers

- *Past*: Necessarily, whenever x is normal at t , there is a true history-proposition p such that p would still have been true if x had blinked at t .
- *Laws*: Necessarily, whenever x is normal at t and p is a true law of nature, p would still have been true if x had blinked at t .
- *Closure*: Necessarily, whenever x is normal at t and p is metaphysically necessitated by a set of propositions each of which would have been true if x had blinked at t , p would have been true if x had blinked at t .

This lead to the unfortunate conclusion that:

- *Triviality*: Necessarily, whenever x is normal at t and determinism is true, each true proposition would still have been true if x had blinked at t .

Closure seems beyond doubt. Dorr argues against *Past*.

Though see Gallow's recent paper.

Lewis's insistence on *Past* is motivated by a genuine insight. When we evaluate an ordinary counterfactual whose antecedent is concerned with a particular interval of time, there is usually some time t —typically only shortly before that interval—such that we tacitly "hold fixed" a very broad range of propositions about history before t , in the sense that we assume that if these propositions are in fact true, they would also have been true in the relevant counterfactual circumstances. When we take ourselves to know one of these propositions, we also treat it as a resource we can draw on in our reasoning about how things would have been different if the antecedent were true. This is how we end up with judgments like the following:

- (13) If John had forgotten to have breakfast this morning, that would have been the first time that he did so in months.
- (14) If I had been honest during the interview, my colleagues would know that I was fired by my previous employer.
- (15) If we convinced a million more people to download this video, we would set a new record.

But this insight can be accommodated by maintaining that we hold fixed *macro-facts*. Can we?

there is no guarantee whatever that [a world where the actual laws are true and where Nixon presses the button] can be chosen so that the differences diminish and eventually become negligible in the more and more remote past. Indeed, it is hard to imagine how two deterministic worlds anything like ours could possibly remain just a little bit different for very long. There are altogether too many opportunities for little differences to give rise to bigger differences. Lewis

Our best deterministic physical theories have continuous dynamics, which means that so long as the past is not infinite, we can always find a nomically possible world which stays arbitrarily close to the actual world throughout any finite initial segment of history, just by choosing an initial state that is close enough to that of the actual world. . . . Could getting t into the particular region of the state space it needs to occupy for Nixon to press the button require a trajectory that diverges substantially from actuality long before t , so that many ordinary sentences about history before t get different truth values? In principle the answer could be yes, but it is extremely unlikely. The key to seeing why is the fact emphasised by Lewis, that little differences characteristically blow up quickly into much bigger differences. . . . the macropresent screens off the macrofuture from the macropast. . . . in the probability distribution that we get by restricting the natural volume measure to a particular macrostate, facts about future macrostates are, approximately, probabilistically independent of facts about past macrostates.

Worries:

- (16) If a big comet had hit Washington, D.C. yesterday afternoon, the U.S.A. would have been left without a President.
- (17) If we had aimed the electron microscope a tenth of a degree further to the left, the image of that gold atom would have appeared in the centre of the screen.

So what about backwards causation? We simply have to deny that counterfactual dependence = causation.

4.3 Against denying *Laws*

Frank is a physicist devoted to discovering a mistake.

- (18) If we had given Frank a glass of water, his whole career would have been devoted to a mistake.

Well, perhaps unconvincing, since in that case, his career would have been devoted to something slightly different. (I.e. there's a false *de dicto* reading of this.)

Suppose that Frank greatly values not having devoted his career to a mistake. If regret is constrained by beliefs about counterfactuals. . . then if Frank comes to believe that his career would have been devoted to a mistake if he had not acted in a certain way, he will be unable to rationally, unequivocally regret acting in that way (assuming nothing else of comparable importance is at stake). In that case, coming to believe [\neg *Laws*] will give Frank a sovereign remedy against regret! Whenever he does anything, no matter how foolish, he can immediately afterwards look back and think, 'If I hadn't done that thing, my whole career would have been devoted to a mistake!'. If he really believes this, and no other comparably weighty values are in play, he should be, all things considered, glad that he did the foolish thing. . . . [Likewise] the action with the highest counterfactual expected utility for Frank will simply be whichever action he is most confident he will in fact perform.

we might regard the question whether the laws or the initial conditions are counterfactually robust as contingent, holding that the laws are robust at possible worlds where the initial conditions are much more complex than the laws, while the initial conditions are robust at worlds where the laws are much more complex than the initial conditions. This is in fact the option I would favour, assuming that there are any possible worlds of the latter sort.

4.4 Applied to ability

Dorr observes that essentially the same situation arises in other parts of the literature, in particular about freedom:

- *Past*: Necessarily, whenever x is normal at t , there is a true history-proposition p such that p would still have been true if x had done any of the things x can do at t .
- *Laws*: Necessarily, whenever x is normal at t and p is a true law of nature, p would still have been true if x had done any of the things x can do at t .
- *Closure*: Necessarily, whenever x is normal at t and p is metaphysically necessitated by a set of propositions each of which would still have been true if x had done any of the things x can do at t , p would still have been true if x had done any of the things x can do at t .

This together entail:

- *Triviality*: Necessarily, if x is normal at t and determinism is true, every true proposition p is such that p would still have been true if x had done any of the things x can do at t .

Dorr notes that it's a weird situation if you accept the conclusion of this argument while denying the conclusion of the corresponding argument about counterfactuals. This seems true *even in the absence of a CA*, since the reasons to reject premises in the first argument seem to apply here as well; especially if we reject *Past*.

4.5 It depends on context

Teitel and Holguín: essentially, denying the existence of a standard context.

One natural way to understand the question of whether q would have been the case had p been the case is as a question about what sort of possible pasts would have made p most plausible, give or take our knowledge of the actual facts. This way of understanding the question tends to induce backtracking readings. But there are other, less past-centric ways to understand the question of whether q would have been the case had p been the case. A neglected fact is that this can sometimes be understood as a question about what sorts of laws would have made p most plausible, give or take our knowledge of the actual facts.

- (19) a. Had Jesus performed most of the feats attributed to him in the Bible, he would have had to have had magical powers.
- b. Had Jesus performed most of the feats attributed to him in the Bible, it would have been because of a series of extremely improbable but nonetheless nomically possible fluke events.
- (20) a. Were Michelson and Morley to have measured a significant difference in the speed of light traveling in the direction of the presumed luminiferous aether versus light traveling orthogonal to that direction, it would have been because there really was a luminiferous aether.
- b. Were Michelson and Morley to have measured a significant difference in the speed of light traveling in the direction of the presumed luminiferous aether versus light traveling orthogonal to that direction, it would have been because they made an experimental error somewhere.
- (21) If a pitcher were to throw a baseball at $0.9c$, it would cause a massive nuclear explosion.
- (22) If upon entering this room Nancy had pointed a wand at Frank and proclaimed 'Now is the perfect time to reveal to you that I'm a witch with magical powers—abracadabra!', at which point Frank had gone flying about the room in accordance with the motion of Nancy's wand, then Frank's whole career would have been devoted to a mistake.

'even when we're interested in offering physically realistic answers to what-if questions, we treat possibilities in which small, isolated violations of the laws of physics (that is, miracles) get the ball going to $0.9c$ as more relevant to the assessment of the counterfactual than possibilities in which the ball accelerates to $0.9c$ by nomically respectable means'

References

- Dorr, C. (2016). Against counterfactual miracles. *Philosophical Review*, 125(2):241–286.
 Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
 Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*.