1-1-2017

Causation, Probability, and the Continuity Bind

Anthony F. Peressini
Marquette University, anthony.peressini@marquette.edu

Causation, Probability, and the Continuity Bind

Anthony F. Peressini

December 30, 2015
[Version 3.61w]

Abstract

Analyses of singular (token-level) causation often make use of the idea that a cause increases the probability of its effect. Of particular salience in such accounts are the values of the probability function of the effect, conditional on the presence and absence of the putative cause, analyzed around the times of the events in question: causes are characterized by the effect's probability function being greater when conditionalized upon them. Put this way it becomes clearer that the 'behavior' (continuity) of probability functions in small intervals about the times in question ought to be of concern. In this paper I make an extended case that causal theorists employing the 'probability raising' idea should pay attention to the continuity question. Specifically, if the probability functions are 'jumping about' in ways typical of discontinuous functions, then the stability of the relevant probability increase is called into question. The rub, however, is that sweeping requirements for either continuity or discontinuity are problematic, and as I argue, this constitutes a 'continuity bind'. Hence more subtle considerations and constraints are needed, two of which I consider: (1) utilizing discontinuous first derivatives of continuous probability functions, and (2) abandoning point probability for imprecise (interval) probability.

1   Introduction

Analyses of singular (token-level) causation often make use of the idea that a cause increases the probability of its effect. Of particular salience in such accounts are the values of the probability function of the effect, conditional on the presence and absence of the putative cause, analyzed around the times of the events in question: causes are characterized by the effect's probability function being greater when conditionalized upon them. Put this way it becomes clearer that the 'behavior' (continuity) of probability functions in small temporal intervals about the times in question ought to be of concern. One prominent (but under-examined) account of token-level causation, that of Ellery Eells ([1991]), actually requires point 'jumps' (discontinuities) in the relevant probability functions for positive and negative token-level causes.

In this paper I make an extended case that causal theorists employing the 'probability raising' idea should pay attention to the continuity question, as it has serious implications for the viability of their accounts. Specifically, if the probability functions are 'jumping about' in ways typical of discontinuous functions, then the stability of the relevant probability increase is called into question. The rub, however, is that sweeping requirements for either continuity or discontinuity are problematic, and as I argue, this constitutes a 'continuity bind'. Hence more subtle considerations and constraints are needed.

I begin by introducing the question of continuity in the context of causation and probability functions (trajectories) using the work of Eells—one of the few theorists to explicitly consider continuity.[1] I then show how discontinuity requirements like Eells' are untenable, and in a surprisingly decisive way. Next I argue that the discontinuity issue also has problematic implications for causal accounts without explicit discontinuity requirements (Menzies [1989]; Noordhof [1999]; Hitchcock [2004]; Kvart [2004]; Northcott [2010]; Glynn [2011]). After that I consider a blanket continuity requirement and show that it is, while a tempting response, unworkable because of the need to allow for the possibility of (empirically motivated) discontinuity in probability trajectories. And therein lies the continuity bind. Finally I consider two potential ways out of the bind, including (1) utilizing discontinuous first derivatives of continuous probability trajectories for theoretically motivated discontinuity needs, and (2) abandoning point probability trajectories altogether in favor of imprecise (interval) probability trajectories.

2   Probability Trajectories and Continuity

I present the continuity question here in a causal setting, roughly following (Eells [1991]); I then expand it to other prominent accounts below in Section 3. Let x and y denote token events, where x takes place at time and place (t_x, s_x) and y takes place at (t_y, s_y). Assume that x's being X caused (in some plausible way) y's being Y, where x is of type X and y is of type Y. Of interest is how the probability of token event y's being Y evolves between t_x and t_y, that is, how the probability of y's being Y changes as a function of time. (I will abbreviate the token events of 'x being X' and 'y being Y' by just writing the properties exemplified, X and Y.)

In the ensuing discussion probability will be understood as objective and physical, that is, as a 'physical probabilities' or 'chance'. A single-case time-dependent probability function P will be assumed as part of a probability space triple $< \Omega, \mathscr{F}, P >$, where Ω is a set, $\mathscr{F}$ is a σ-field over Ω, and P is a probability function on $\mathscr{F}$ that obeys the standard (Kolmogorov) axioms of the probability calculus. Physical probabilities apply to particular events, ones that occur

---

[1]Peter Menzies ([1989]) and Igal Kvart ([2004]) in their respective accounts are also sensitive to how the probabilities evolve through time, but without explicitly addressing continuity one way or the other; I take this up further below.

or fail to occur at a particular time and place, and hence have values defined relative to a time of evaluation. To make explicit the temporal index, t, involved in evaluating the probability of event $Y \in \mathscr{F}$ at time t, I use the notation $P_Y(t)$. In general if an event Y occurs at a time t_y, $P_Y(t)$ is strictly between 0 and 1 prior to t_y, and 1 at time t_y and all later times. Following (Eells [1991]) I use the term 'probability trajectory' to refer to a probability function understood in this way as a function of time.[2]

Again, while there are challenges to any interpretation of probability, in this causal setting an objective physical understanding akin to 'chance' is a reasonable way to proceed; see (Eells [1991], pp. 34–55) and (Eells [2010])) for detailed discussions. I (loosely) follow Jenann Ismael ([2011]), pp. 419–20) with my pre-theoretic understanding of physical probability or chance, taking it to be '…the link between the fundamental level of physical description in quantum mechanics and the measurement results that mark the points of empirical contact between theory and world'. I follow her in that my understanding is that physical probability is objective and non-trivial (not everywhere zero or one). I remain agnostic, however, with respect to her ultimate analysis of it—especially whether its grounding is at the quantum level or some higher level as in (Glynn [2010],[forthcoming]), or (Sober [2010]).

2.1 Probability trajectories

As an illustration of a probability trajectory in action, consider the following example from Deborah Rosen [1978], modified here from (Eells [1991]):

Example 1 A poorly putted golf ball is rolling roughly in the direction of the cup when a squirrel runs by and bumps it in such a way that its resulting trajectory is directly toward the cup and it continues right into the cup.

Following the standard assumptions of such causal discussions, I take the probability values of $P_Y(t)$ to reflect the objective probability of the event (ball going in the hole) and assume that it is strictly less than one until Y occurs. Suppose that the probability of the ball going into the cup given its initial trajectory, velocity, and so on, is 0.25. Suppose further that, in general, the (type) probability of balls going in when squirrels bump them is very low (say 0.05), however, in this (token) case the particular trajectory of the ball immediately following the bump was such that the probability of the ball falling in the cup was rather high, say 0.8. Let the event of the squirrel bumping the ball be x being X and the event of the ball going into the cup be y being Y. The probability trajectory of Y can be depicted following (Eells [1991], p. 293) as in Figure 1.

The standard analysis of this example is that the squirrel's kick X caused the ball to drop into the cup Y, despite the fact that, in general, squirrel kick's in such situations almost never result in the ball going in the hole. For causal considerations, the salient features of the graph are that the probability of Y takes an immediate point drop in probability at t_x, corresponding to the type-level fact that X-type events generally decrease the probability of Y-type events, and that the probability of Y recovers immediately after the ball is bumped at t_x to a higher value than it had before because of the favorable trajectory and velocity actually imparted by the token event X. Hopefully this causal story is plausible enough, though its causal details are not the primary

---

[2]For this paper, I will understand the basic form of these probabilities as unconditional. This is distinct from general probability, which applies to classes of event and whose basic forms is conditional. This for clarity and convenience only: the continuity issues I deal with here are not sensitive to whether the probabilities are analyzed in the standard Kolmogorovian way or some other way, with a different conditionalization rule and/or with conditional probabilities as the basic form; see (Hájek [2003]).
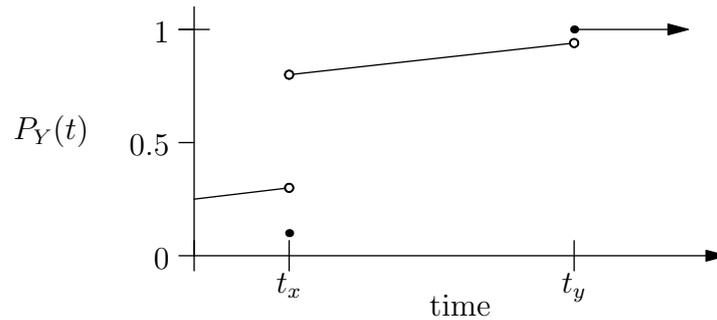
Figure 1: Probability trajectory with discontinuous jump at occurring event.

concern here. For present purposes, the crucial features of the graph are the discontinuities at t_x and t_y, that is, the fact that the probability of Y 'jumps' up just after x happens and then 'jumps' again to 1 at moment the ball falls into the cup.

As can be seen from the graph, Eells employs jump discontinuities in the probability trajectory to (a) indicate that a token (positive) 'cause' has taken place, and (b) emphasize that the world is chancy or indeterministic at the macro-level in that the probability of the event in question is bound away from 1 until it happens.[3] Use (a) is central to his account (as I detail below), I will call this general idea the

CDJP (Causal Discontinuous Jump Principle) The probability trajectory of an event e that occurs at t_e jumps discontinuously at times when events causally relevant to e occur.

The status of the second use of discontinuity, (b), is less clear in Eells work. This 'occurring event discontinuity' assumption is made and discussed explicitly (Eells [1991], p. 294). I will refer to this assumption as the

DJP (Discontinuous Jump Principle) The probability trajectory of an event e that occurs at t_e jumps discontinuously to 1 at time t_e.

Notice though that the assumption that the probability of an event is not one until the event occurs is also consistent with the graph continuously approaching one from below. Eells recognizes that the indeterminism could also be represented in a continuous fashion, with the probability continuously approaching one from below. But he writes that his analysis does not 'pay attention' to whether the trajectory is continuous at the time the event occurs (Eells [1991], p. 294, note 6]), and does not explicitly committed himself to (DJP), though he does consistently draw all his graphs with such a discontinuity. Consider the alternative graph depicted in Figure 2 in which the probability trajectory continuously approaches one at t_y. It is equally true in this graph that the probability of Y is strictly less than one until it actually occurs at t_y. The difference between this graph and the graph in Figure 1 is that in Figure 1 the value of $P_Y(t)$ is bound away from one prior to t_y, while in Figure 2 the value of $P_Y(t)$ becomes arbitrarily close to, but always less than one as t approaches t_y. I argue below that Eells was mistaken about nothing turning on DJP and that thinkers concerned with probability should indeed 'pay attention' to this continuity issue. But first I sketch his causal account.

---

[3]A 'jump' discontinuity is one in which the left- and right-hand limits exist, but are not equal. It is important to note that in order for there to be a discontinuous jump (jump discontinuity) as Y occurs (or at any other significant time, for example, t_x), it is necessary that the trajectory be continuous in some (perhaps very small) interval to the left of the jump discontinuity—this will become significant below. The other two possibilities, that the left and right hand limits exist and are equal, or that one (or both) fail to exist are called 'removable' and 'essential' discontinuities, respectively. The essential discontinuity case will come up again below.
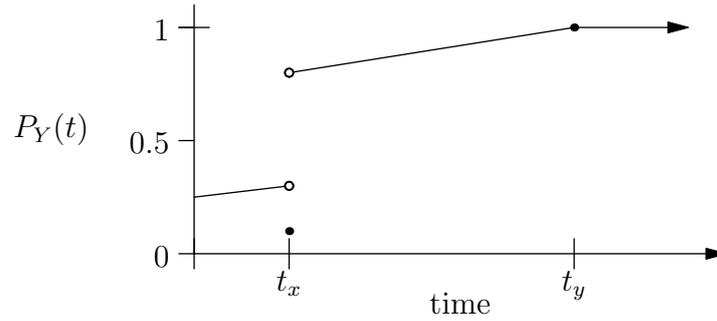
Figure 2: Probability trajectory with continuous probability at occurring event.

2.2 Causation as discontinuous jumps

In his treatment of this example, Eells depicts the probability trajectory of $P_Y(t)$ as in the graph in Figure 1. The crucial feature of this graph for the causal question is that the probability of the ball falling in the cup is higher immediately after the ball is bumped at t_x than it was prior to the bump, despite the fact that in general such types of events lower the probability of the ball going in as indicated in the graph by the immediate point drop in probability at t_x.[4] According to Eells, this structural property of the example is what leads us to say that this token squirrel bump caused the ball to go in, despite the fact that, in general, such bumps tend to prevent balls going in rather than cause them. When the probability trajectory of an event y has this structure Eells defines y to have occurred 'because' x occurred. Explicitly what is required for an event y's being Y 'because' event x was X is the following three conditions:

  (i) the probability of Y changes at the time of x,

 (ii) immediately after the time of x the probability of Y is both high and higher than before x, and

(iii) this probability remains high until the time of y.

Eells describes three additional causal relations: that of an event's occurring 'despite' another event, and events being 'independent' and 'autonomous' of another event. Though the details of the additional three will not be of particular concern here, the basic idea is that in the despite case, the probability decreases (and remains low), in the independent case it remains the same, and in the autonomous case the probability increases to a high level but then drops to a low level. Eells ([1991], p. 355) defines each of these relations in terms of the left and rights limits of $P_Y(t)$ at t_x. He also specifies the qualifications needed to preserve the 'causes increase the probability of their effects' idea, in particular that one must hold fixed the set, K, of the actual, separate, independent causes of Y and also any interactive factors by which X influences the probability of Y (Eells [1991], Sec. 6.4). In order to build in the temporal evolution of the probability in the proper way, Eells further specifies W_t be the conjunction of all factors of the world (relevant to y's being Y) that have fallen into place by time t and whose exemplification (relative to K) can be traced back to the exemplification of X at t_x. He then defines $P_Y(t) = P(Y|K\&W_t)$ for all times t. Against this understanding of the probability trajectory, the degree to which y is Y

---

[4]Another way of putting this, in terms of type-level probabilistic causality, is that the bump is a token-level positive causal factor for the ball going in but that bumps of this type are type-level negative causal factors.

because of, despite, independently of, and autonomously x's being X are defined as follows:

$$
\begin{aligned}
B &= M - P^- \\
D &= P^- - P^+ \\
I &= 1 - |P^- - P^+| \\
A &= P^+ - M
\end{aligned}
$$

where P^- and P^+ are the left and right hand limits of $P_Y(t)$ at t_x, respectively, and $M = \min\{P_Y(t) : t \text{ in } (t_x, t_y)\}$, which can be thought of intuitively as the lowest point in the probability trajectory between t_x and t_y. It should be clear that in the case of 'despite', if X is to play a negative ('despite') causal role ($D > 0$), that $P^- - P^+ > 0$, from which it follows that $P^+ \neq P^-$, or in other words that the limits of $P_Y(t)$ from the left and the right at t_x are not equal, which entails a jump discontinuity at t_x. The same follows for the 'because' relationship as well.[5] Thus if X is a positive or negative token cause of Y, then $P_Y(t)$ has a jump discontinuity at the time of X. That is, Eells' account requires an event's probability trajectory to have discontinuous jumps at times at which causally relevant events occur.[6]

Let me stress again that Eells' official position is that it is inconsequential as to whether occurring events jump discontinuously to one (DJP), though he favored and utilized the discontinuous version. But as to the question of whether causes require a jump discontinuity in the probability trajectory of their effects (CDJP), Eell's answer is an unequivocal 'yes', since his entire token-level account depends on such discontinuities. It is surprising that this aspect of his influential account has received virtually no discussion.

2.3   Against systematic discontinuity

In this section I make a detailed case for why systematic discontinuity requirements like DJP or CDJP are problematic. I first direct my case against DJP. The reasons for beginning with DJP are: (1) DJP is more general and is therefore of interest in its own right, outside the setting of probabilistic causality (for example, event ontology, chance, and so on), and (2) the argument is more straightforward and perspicuous in the case of DJP and requires only minor adjustment to apply to CDJP as well. In what follows I present the main thread of a formal argument against DJP (the details of which can be found Appendix A) and then show how it extends to CDJP.

Reconsider Example 1—especially the period from after the time the squirrel bumps the ball to the time it enters the cup. The instant the ball comes off the bump it has a certain trajectory and speed, one that will take it directly into the cup; this is why the probability of Y is high after that instant. As time gets closer to t_y and the ball gets closer to the cup, the number of eventualities that could prevent the fall into the cup decreases, and so its probability continues to increase. In other words, as the ball passes by points on the green closer and closer to the cup and with the same favorable trajectory and speed, the probability of its going in the cup would naturally be expected to continue to get closer and closer to one. While these considerations alone favor a continuous increase of $P_Y(t)$ to one, a stronger case can be made.

---

[5]This is because by definition, $M \leq P^+$, so if $B = M - P^- > 0$, then $P^+ - P^- > 0$, so $P^+ \neq P^-$, that is, the limits of $P_Y(t)$ from the left and the right at t_x are not equal, which again entails a jump discontinuity at t_x.

[6]Eells ([1991], p. 354) makes it explicit that it is a presupposition of his account that the left and right limits of $P_Y(t)$ at t_x exist. The value at the point t_x itself is not constrained. At another point, Eells ([1991], p. 355) also requires that the probability trajectory actually be constant in some open interval both to the left and right of t_x. He may have sensed the possibility of pathological behavior around such discontinuities. In any event, I argue below that his required 'jump' discontinuity will entail not only that $P_Y(t)$ cannot be constant to the left of t_x, but that it cannot even have a limit from the left.

If the probability trajectories of all occurring events jump discontinuously at the instant they occur (DJP), then in particular the probability trajectory for each of the occurring events 'leading up' (causally) to the event under consideration would also have a jump discontinuity at the time each of them occur. The probability (trajectory) of the original event is not independent of the probability (trajectories) of certain events leading up to it, that is, its probability depends on certain events that need to 'fall into place' in order for it to happen—and this gives rise to problems. Returning to Example 1: between the time of the cause t_x and the time of the effect t_y, both version's graphs (Figures 1 and 2) depict the probability trajectory as continuous in the interval just to the left of t_y. This, however, does not accord with the 'jumpy' nature of the probabilistically relevant prior events falling into place. If all the events involved in the ball traversing the points on the green after being bumped and before entering the cup have probability trajectories that have a jump discontinuity at the time they occur, then it seems that the probability trajectory of Y (ball falling into the cup), which depends upon these events falling into place, should reflect this discontinuous 'jumping' at the times these prior events occur.

Such considerations suggest that the discontinuous jump (as mandated by DJP) in the probability trajectory of an occurring event is inconsistent with the probability trajectory $P_Y(t)$ being continuous in the interval just before t_y, as it must be in order to have a jump discontinuity. If this is right, then assuming DJP in such settings is inconsistent, since $P_Y(t)$ is required (as depicted) to be continuous in at least some small interval to the left of t_y. I now put this objection on a formal footing to show more precisely the source of the problem.

The form the argument will take is that of a inconsistent/incoherent dilemma, namely that DJP in this setting entails either that

1. the probability trajectory, $P_Y(t)$, is discontinuous from the left at t_y (has no left hand limit), which is inconsistent with there being a jump discontinuity at t_y, or that

2. the certainty (distance the probability is from 1) of antecedent events upon which Y depends becomes arbitrarily larger than the certainty of Y itself, which will be shown to be an incoherent result.

For definiteness, the setting will parallel Example 1 and concern the assessment of the causal relevance of a token event x being X for event y being Y, where these events occur at t_x and t_y, respectively. I will show that DJP entails the unintended (and unexpected) consequence that $P_Y(t)$ has no limit from the left (is left discontinuous) at t_y. To get the argument off the ground I make use of a well-known theorem from probability, Bayes' Theorem, which states that:

$$P_{Y|X}(t) = \frac{P_{X|Y}(t)P_Y(t)}{P_X(t)}, \quad P_X(t) > 0$$

where $P_{Y|X}(t)$ is the probability of Y conditional on X at time t, and similarly for $P_{X|Y}(t)$. A simple variation of this that will be useful here is:

$$P_Y(t) = \frac{P_{Y|X}(t)P_X(t)}{P_{X|Y}(t)}, \quad P_{X|Y}(t) > 0 \tag{2.1}$$

The role of Equation (2.1) will be to instantiate in a formal way the intuitive idea expressed above by the idea that the probability (trajectory) of the event under consideration depends somehow on the probability (trajectories) of the events that fall into place leading up to it.[7]

---

[7]Recall that the causal background context K is built into the definition in that $P_Y(t) = P(Y|K\&W_t)$, so for

Consider a sequence of moments $\{\hat{t}_i\}$ converging to t_y (the moment Y occurs) and a sequence of events $\{X_i\}$ occurring at these times and upon which Y probabilistically depends. In the context of Example 1, these events and moments will be where the ball was at half of a second before it went in, and a fourth of a second, an eighth of a second, and so on. More formally we might put this as $\hat{t}_i = t_y - \frac{1}{2^i}$ and X_i = the event of the ball being where it was at $\hat{t}_i$ with the particular favorable trajectory it had.[8] In order to formalize the degree to which probability trajectories jump discontinuously to one, I use the left-hand limit of the probability of an event X as time approaches the time X occurs (from before) and define the 'distance' of the jump to one as:

$$D(X,t) = \lim_{s \to t^-} 1 - P_X(s) \tag{2.2}$$

The value of $D(X, t_x)$, where t_x is the time X occurs, is the distance the probability trajectory jumps to reach one when X occurs. That $D(X, t_x)$ is greater than zero for all events is equivalent to the assumption of DJP.

Returning to the sequence $\{X_i\}$, we see that $\{D(X_i, \hat{t}_i)\}$ might be thought of as the 'degree of indeterminism' or "chanciness" for each of the X_i events. In terms of Example 1, it would be values representing how far the probability of the ball being where it is at a half of second before it goes in, a fourth of a second, an eighth of a second, . . . jumps as each of those events occur. The $\{X_i\}$ sequence of events leading up to Y will be used to show that DJP is problematic. At this point the argument bifurcates based on whether the sequence $\{D(X_i, \hat{t}_i)\}$ converges to zero, that is, whether

$$\lim_{i \to \infty} D(X_i, \hat{t}_i) = 0. \tag{2.3}$$

I will show that if it does not converge to zero, then we have the inconsistency horn of the dilemma, and if it does converge to zero we have the incoherency horn.

2.3.1 Inconsistency

The general strategy for this horn is to show how, using Bayes' Theorem, if the 'event occurring jumps' of the events getting close in time to Y do not converge to zero, then this forces the probability trajectory $P_Y(t)$ to be discontinuous to the left of t_y (not have a limit as time approaches t_y from the left), which contradicts the requirement that the left hand limit of $P_Y(t)$ at t_y exists. The full details of the proof can be found in Appendix A, which I sketch below.

Let $0 < L < 1$ be the left hand limit of $P_Y(t)$ at the time Y occurs, t_y. To show that $P_Y(t)$ is discontinuous from the left at t_y it is sufficient to find a sequence of times $\{t_i\}$ converging to t_y

---

example, $P_{Y|X}(t)$ will be $P(Y|X\&K\&W_t)$. In the case of the example under consideration here, where the X's are not arbitrary events, but rather ones causally relevant to Y and converging to Y's time, all the probability functions in what follows will have the same causal background context; I will generally suppress the addition notation in what follows, noting it only where it makes a difference.

[8]As will become clear below, a key feature of the X_i's is that they be at least probabilistically and/or causally relevant to Y. Thus the construction is unproblematic when there is a space-time process leading up to or constituting the event Y as in the case here. The argument does not, however, necessarily apply to quantum events or certain kinds of macro-level events that embed quantum event, which may not under certain interpretations have such probabilistically relevant antecedent events. (I will not get into the distinction between causally and probabilistically relevant here; there is much debate about when and how these notions coincide. In most settings probabilistic relevance is weaker and I will employ it here.) While there is debate about whether all macro-level examples of causation need to have such an intermediate process, even accepting a pluralistic view (Hall [2004]), for my argument here it is sufficient that it work for the large class of macro-level cases like Example 1 in which there is such a mediating process. See note 10 below for more on the kinds of quantum-embedding macro events that are excluded.

such that $P_Y(t)$ evaluated at those points does not converge to L, that is, $\{P_Y(t_i)\} \not\to L$. In terms of the definition of convergence this means showing that there is an $\epsilon > 0$ such that

$$|P_Y(t_i) - L| \geq \epsilon, \quad \text{for some } i \text{ greater than any } N > 0. \tag{2.4}$$

The next task is to actually construct the problematic sequence $\{t_i\}$ and show that it satisfies Equation (2.4).

Since by hypothesis $\{D(X_i, \hat{t}_i)\}$ does not converge to zero, there is an $\hat{\epsilon} > 0$ such that

$$D(X_i, \hat{t}_i) \geq \hat{\epsilon}, \quad \text{for some } i \text{ greater than any } N > 0. \tag{2.5}$$

It will simplify notation to define $\{L_i\}$ to be the sequence of left hand limits for each of the $\{X_i\}$, so $L_i = 1 - D(X_i, \hat{t}_i)$, from which it follows that $|1 - L_i| \geq \hat{\epsilon}$ because of Equation (2.5). To construct the sequence of times, $\{t_i\}$ that will generate the contradiction, we must find a sequence of moments slightly before each of the $\{\hat{t}_i\}$, since we will be interested in what is happening to the probabilities of the $\{X_i\}$'s right before the time they occur—recall that $P_{X_i}(\hat{t}_i)$ is simply equal to 1. A natural choice would be to pick moments like the following: $\hat{t}_i - \frac{1}{10^i}$. Such a sequence of moments are always just before the moment $\{X_i\}$ occurs and are such that as $i \to \infty$, they get arbitrarily close to those moments, $\hat{t}_i$. But we need the new sequence to be more "tightly" tied to $P_{X_i}(t)$.

The factor required will depend on each of the $P_{X_i}(t)$ and how quickly each approach their limit near $\hat{t}_i$. The limit in question is L_i, so by the definition of limit, we can find a minimal distance $\delta > 0$ such that the distance between $P_{X_i}(t)$ and L_i can be made less than an arbitrary $\epsilon > 0$ for all t within the minimal distance δ of $\hat{t}_i$. The ϵ we will use here comes from the assumption that $D(X_i, \hat{t}_i) \not\to 0$, namely, $\hat{\epsilon}/2$, which is half of the $\hat{\epsilon} > 0$ we have from Equation 2.5. Now consider the sequence of $\{\delta_i\}$, each greater than zero, with the property that if $|\hat{t}_i - t| < \delta_i$, then $|L_i - P_{X_i}(t)| < \hat{\epsilon}/2$. Again, that the $\delta_i > 0$ exist follows from the definition of the (left) limit of $P_{X_i}(t)$ at $\hat{t}_i$ being L_i. Finally we define the $\{t_i\}$ as: $t_i = \hat{t}_i - \frac{\delta_i}{10^i}$ Thus the sequence of $\{t_i\}$ are such that as $i \to \infty$, the t_i's get arbitrarily close to the $\hat{t}_i$'s (the times that the X_i's occur), and further, each t_i is within δ_i of $\hat{t}_i$, so $|L_i - P_{X_i}(t_i)| < \hat{\epsilon}/2$, for each t_i. From which it follows that

$$1 - P_{X_i}(t_i) \geq \epsilon, \quad \text{for some } i \text{ greater than any } N > 0, \tag{2.6}$$

where $\epsilon = \hat{\epsilon}/2$. (See Appendix A, near Equation (A.3) for details.)

In making use of Bayes' Theorem, the conditional probabilities $P_{YX_i}(t_i)$ and $P_{X_iY}(t_i)$ will be needed. It will be assumed that $\lim_{i\to\infty} P_{YX_i}(t_i) = L$ and $\lim_{i\to\infty} P_{X_iY}(t_i) = 1$. The reasoning for this is as follows: given that the limit of $P_Y(t_i)$ is L, conditionalizing on the X_i's, which are the particular events leading up to Y, should not affect convergence; similarly since (by definition) the $\{X_i\}$ are events leading up to Y at times $\{t_i\}$ converging to the time they occur $\hat{t}_i$, the probability of these events at these times conditionalized on Y will naturally converge to 1 as time converges to the time of Y.[9] The convergence of the conditional sequence $P_{YX_i}(t_i)$ to L entails that for any $\epsilon_1 > 0$ there is an $N_1 > 0$ such that for all $i > N_1$

$$|P_{YX_i}(t_i) - L| < \epsilon_1 \tag{2.7}$$

Similarly, the convergence of the conditional sequence $P_{X_iY}(t_i)$ to 1 entails that for any $\epsilon_2 > 0$ there is an $N_2 > 0$ such that for all $i > N_2$

$$|1 - P_{X_iY}(t_i)| < \epsilon_2 \tag{2.8}$$

---

[9]Nothing turns on these particular values for the limits of the conditional probabilities. As long as they converge to some $0 < \hat{L} \leq 1$, as they must, then the proof can proceed with simple scaler adjustments.

From this appropriate values can be chosen for ϵ_1 and ϵ_2 in terms the ϵ from Equation (2.6) and then Equation (2.1), Bayes' Theorem, can be utilized to show that

$$|P_Y(t_i) - L| \geq \epsilon' > 0, \tag{2.9}$$

for some i greater than any $N > \max\{N_1, N_2\} > 0$, where $\epsilon' = \frac{\epsilon L(\epsilon+4)}{4(\epsilon+2)}$. Thus $P_Y(t)$ is discontinuous from the left at t_y. (See Appendix A, Equations (A.7) through (A.11) for details.)

2.3.2   Incoherence

Having shown that if the sequence of 'indeterministic jumps' $\{D(X_i, \hat{t}_i)\}$ does not converge to zero, then we have the contradictory result that $P_Y(t)$ is discontinuous from the left at t_y, consider now the case in which $\{D(X_i, \hat{t}_i)\} \to 0$.

Again, this means that the difference between the left-hand limit of $P_{X_i}(t)$ at $\hat{t}_i$ and 1 can be made arbitrarily small for large enough i. That is, we can find events X_i probabilistically relevant to Y and arbitrarily close to the time of Y by picking an i large enough such that the probability of X_i just before $\hat{t}_i$ is arbitrarily close to one—this despite the fact that the probability of Y at $\hat{t}_i$ is bound away from 1. For example, this means that the probability of the ball being at a point arbitrarily close to falling into the cup at an arbitrarily small instant before (say ϵ) it actually does is a million, or a billion, or trillion, ... times closer to 1 than the probability of the ball's falling into the cup the same arbitrarily small instant ϵ before it actually does.

More concretely, pick any location and time that the ball was very close to falling in, say within $\frac{1}{10^{256}}$ of a second and $\frac{1}{10^{512}}$ of an inch from the edge. At an even smaller instant before the ball was at this location, its probability of being there was any huge number you like, say 10^{1024}, times closer to 1 than the ball's probability was when it was a trillion (or any huge number you like) times closer to falling in.

In short, this result says that while the probabilistically relevant antecedent events of the ball being closer and closer to the hole with the favorable trajectory it had are such that their probability right before they happen are getting as close to 1 as you like, the probability of the ball falling in the hole, as close to the time it did as you like, is as many times farther away from 1 as you want to make it. That an event's probability right before it happens is arbitrarily farther away from 1 than is each of an infinitesimally close series of (probabilistically) relevant events leading up to it is unintelligible. This I offer as the incoherency horn of the dilemma.[10]

2.3.3   Extending to CDJP

Having made the case that DJP is untenable, it is relatively straightforward to extend this to CDJP. The same construction of the sequence of events $\{X_i\}$ and $\{t_i\}$ can be used with the simple

---

[10]I stress again that the argument of this section is against the view embodied in DJP, namely, that all occurring events must jump discontinuously to 1; it is does not militate against the (very plausible) view that some classes of events might so jump. In particular, it is compatible with the view that the probability trajectories of quantum events or certain kinds of macro-level events that embed quantum events in particular ways do in fact so jump.

As a reviewer for this journal points out, without such a limitation a quantum variation of my Example 1 might be construed as a counterexample to the incoherence horn of the dilemma. Consider Example 1 modified so that the golf ball's fall into the cup triggers a quantum triggering device that has an irreducible chance of 0.9 of triggering an explosion nearby. One must assume (*per impossible*) that the triggering process and the detonation, if they happen, will both be instantaneous, so if the explosion happens, it will happen at the same moment t_y when the ball falls into the hole. Then if the ball being where it was with its favorable trajectory at moments converging to the time it falls into the cup are the events $\{(X_i, \hat{t}_i)\}$, and the explosion is Y, the $\{(X_i, \hat{t}_i)\}$ events are getting arbitrarily close to Y and their probabilities are getting arbitrarily close to one, but the probability of Y is bound away from one (at 0.9). Whether it is legitimate in this context to collapse such triggering events onto Y in the probability trajectory is not clear, but fortunately it need not be taken up here.

change that $\{t_i\}$ converges to t_x (the moment x occurs) instead of t_y, and so X_i = the event of the ball and the squirrel being where they were and in the particular states (velocity, direction, internal states, and so on) at t_i, which is converging to t_x. In the case of DJP, we exploited the fact that $P_Y(t)$ has to 'jump' because of its dependence (via Bayes' Theorem) on the $P_{X_i}(t)$, which jump as X_i occurs at t_i, but in the case of CDJP the problem is more immediate.

It seems incontrovertible that the $\{X_i\}$ events, are positive ('because') token causes, especially for large i when t_i gets very (arbitrarily) close to t_x.[11] But according to CDJP this would require that $P_Y(t)$ must at the very least have jump discontinuities at each of the $\{t_i\}$, corresponding to each of the $\{X_i\}$ being a 'because' token causal factor—and this raises numerous problems. First, Eells explicitly states that it is a requirement that $P_Y(t)$ be constant in some small interval to the left (and right) of t_x, but this cannot be the case if each of the $\{X_i\}$ are 'because' token causal factors. Second, things would seem to get even worse in that there are an uncountable number of possible such 'because' token causal events just prior to t_x, which might well be in tension with core parts of real analysis.[12]

But there is a potential way out, which is related to how Eells deals with suggestions that his account cannot distinguish between simultaneous events as far as causal relevance goes. Recalling that the causal background context K and temporal indexical set W_t is built into the definition of the probability trajectory, that is, $P_Y(t) = P(Y|K\&W_t)$, it might be open to him to deny that the probability trajectories are the same trajectory (function) for each of the $\{X_i\}$ events and the X event.[13] In other words, when considering X's causal significance for Y we have

$$P_Y(t) = P(Y|K\&W_t),$$

but for each X_i's probability trajectory for Y we have

$$P_{Y_i}(t) = P(Y|K_i\&(W_i)_t).$$

Thus the $P_{Y_i}(t)$ could be such that they were (point-wise) converging to $P_Y(t)$ (from below) and still jumping (as required) at the critical time t_i. In particular, the $P_{Y_i}(t)$ functions would presumably be increasing such that $P_{Y_i}(t) \leq P_{Y_{i+1}}(t) \leq \ldots \leq P_Y(t)$, for $t < t_x$. See Figure 3. A virtue of this response would be that the increasing nature of the probability (trajectories) near the critical point t_i would fit neatly with the idea that as each of the events in question, the ball and squirrel being where they were (X_i) at instants closer and closer to when they collide (t_i) in the favorable way they did with respect to causing the ball to fall in the hole (Y), actually occur, the probability of Y increases.

Unfortunately, however, this response is not available on Eells' account as it stands. The first, not insurmountable problem is that as mentioned above, Eells requires that probability trajectories be constant in some open interval to the left (and right) of the jump discontinuities. Eells was (rightly) concerned about the limits upon which his account so crucially turns, and in personal correspondence he indicated that the reason for requiring the trajectories to be constant

---

[11] See (Kvart [2004], p. 369–70) for a similar use of such an example and discussion.

[12] There are provable restrictions on the size and nature of the set of points of discontinuity for real-valued functions. In particular, this set must be an F_σ set, that is, one that can be written as a countable union of closed sets of real numbers (Royden [1988], p. 53). And if the function in question is monotonically increasing or decreasing on an open interval, as a probability trajectory might well be eventually, then there can be no essential discontinuities and at most countably many "jump" discontinuities; see (Rudin [1976], p. 95–7) for details.

[13] This is how Eells' account avoids the putative defect that, for example, an event located far away, say another squirrel kicking a tree, but taking place at the same time as x, would also be deemed a cause of Y because the trajectory would jump at the time of both events, t_x. But in fact, there would be two trajectories, one for X and one for the far away event, and the latter trajectory would not have a 'because' jump because its (distinct) W_t would not include the factors (collision, change in the ball's trajectory, and so on) traceable back only to x being X.
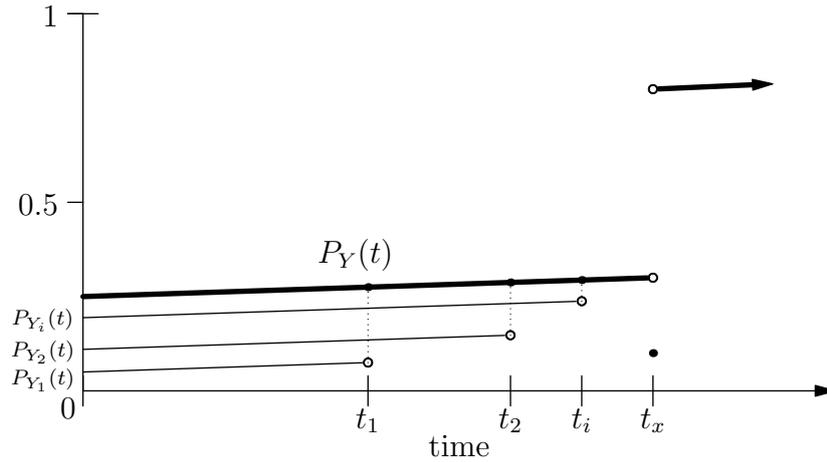
Figure 3: Probability trajectories $P_{Y_i}(t)$ for Y with respect to X_i converging to $P_Y(t)$ from below.

in this way was indeed in part to ensure that the limits exist, which of course it would, but at the steep cost of coherency. To simply require that the left and right hand limits exist, and not necessarily be constant in any open intervals to the left (or right), would seem to be a less restrictive alternative, and it would allow the 'series of increasing trajectories response' I just sketched. But there is a second problem—and that is that the (extensional) way that Eells has defined the causal background context K and temporal indexical set W_t does not allow him to assert that the trajectories associated with the X_i's causal relevance for Y are distinct from X's trajectory, $P_Y(t)$. This is because K is defined to be the set of factors causally independent of X or X_i that are causally relevant to y's being Y in the actual situation, and thus are identical for X and X_i. And as for the W_t, the (actual) factors relevant to y's being Y that are not in K but have fallen into place by time t, these too will coincide in the cases of the X_i and X.[14]

It seems then that utilizing systematic discontinuities for understanding occurring events (DJP) or causation (CDJP) is untenable, but it turns out that even the threat of discontinuities in probability trajectories gives rise to problems for causal theorists, as I take up next.[15]

---

[14]Eells ([1991], pp. 344–5) does allow that one might conceive of K as a 'kind of population', and this could open the possibility of a non-extensional understanding that might allow one to individuate the trajectories in a more fine-grained way. Of course such a move brings with it problems of its own. The account could also potentially be modified in a way that allows for (appropriately bounded) essential discontinuities as well as jump discontinuities. As long as a proper bound was in place, one could employ the 'limit superior' (least upper bound of the cluster points) from the left, rather than a limit from the left. This would potentially alleviate issues with isolated essential discontinuities, though the problematic proliferation of them due to DJP/CDJP would remain as would the problems to be taken up in the next section. I owe thanks to a anonymous reviewer for helping me see this possibility.

[15]As a reviewer points out, another possibility for handling unruly kinds or numbers of discontinuities in probability trajectories would be to move to an account along the lines of (Woodward [2003]) or (Pearl [2009]) that distinguishes between the actual situation of interest and causal models of the situation. Such an approach, in utilizing a model to analyze causation, considers only a fixed and generally discrete set of causal variables, thus even if in the actual situation the probability trajectory had (lots of) essential discontinuities, within the confines of the idealized model such 'bad behavior' might well be modeled sufficiently by a simple jump discontinuity. Indeed, if in the end the continuity bind proves insurmountable, this could be seen as a case for such model based accounts.

3   Broader Discontinuity Concerns

The attention Eells paid to continuity was well placed: how the probabilities evolve through time is both relevant and non-trivial. Such continuity issues ought to be of more concern to all causal theorists exploring probabilistic accounts. Most other prominent accounts have the opposite concern from Eells: they depend—at least implicitly—on the continuity of probability trajectories.

Accounts most obviously embroiled with continuity are those that make explicit use of temporal conditions, for example, Peter Menzies ([1989]), who utilizes 'temporally dense' chains of (counterfactual) probability increases, and Igal Kvart ([2004]), whose account looks for 'stable screeners' and 'causal relevance neutralizers' in temporally intermediate events between cause and effect. If the temporal evolution of the probability values in question cannot be assumed to be continuous, this strains such accounts by rendering the probability in the interval potentially unstable in that it may 'jump' between values that may or may not preserve the presence of the relevant 'probability increases' or the absence of 'stable screeners' (probability decreasers).

Menzies' account does not explicitly address continuity, but it does, however, implicitly constrain discontinuities. Building on David Lewis' ([1986]) counterfactual analysis in terms of unconditional probabilities, Menzies requires that causally related events c and e be 'probabilistically dependent.' This means for him that there must be intermediate events corresponding to any finite set of intervening times between the times of c and e such that the actual probability of each of the intervening events is significantly higher than it would have been had the immediately preceding event in the set not happened. Put in terms of Eells' probability trajectories, this effectively requires the probability function to be monotonically increasing between the times of c and e, and thereby limits the number and kind of possible discontinuities (recall note 12).[16]

Kvart seems to be sensitive to the possibility of the inequality flipping at intermediate temporal points between c and e and offers a condition that may be intended to prevent it. The condition is developed in an example in a section entitled 'Illustrating Causal Relevance through Infinite Regress' (Kvart [2004], Sec. 6, pp. 369–70), but is difficult to assess. The example involves considering a series of events at intervening moments between putative cause and effect and verifying that the inequality does not reverse at any of the points—that they are not 'neutralizers'. Causal relevance is defeated if one of the moments reverses the inequality. But if one can continue indefinitely without a reversal, one thereby constructs an infinite series of moments that are not neutralizers. His full condition is that 'there is only infinite regress of this sort, (i.e., there not being a suitable terminating chain)' and this establishes that c is causally relevant to e, since it guarantees that 'there is no neutralizer for c and e' (Kvart [2004], p. 370). The logical form of the condition seems to be that all such sequences of intervening events converging to an intervening time must not terminate with (contain) a neutralizer. If so, then it might be understood as employing (part of) an alternative specification of the standard epsilon-delta definition of continuity—constraining all convergent sequences in the domain—but in such a way that it entails not full continuity, but rather something close to Menzies' monotonically increasing condition.

While other prominent probabilistic accounts of causation may eschew explicit temporal conditions, not surprisingly, they are not able avoid temporal (and hence continuity) issues

---

[16]This turns out to be an implausibly strong condition and Menzies ([1996]) himself disavows even an amended version of this theory. This tension between stability in the probability trajectory and cripplingly strong constraints on it is endemic to the point probability framework, as I take up further below.

altogether. For example, in (Noordhof [1999]), (Hitchcock [2004]), (Northcott [2010]), and (Glynn [2011]) one finds reference to probability inequalities assessed 'shortly before' the time of the cause and/or effect.[17] In some variation these accounts all consider the probability of an event *e* 'just before' it occurs, conditional on the presence and absence of a putative cause *c*. The critical inequalities involve conditional probabilities at moments 'just before' the time of the cause t_c–ϵ or 'shortly before' the time of effect t_e–ϵ, and perhaps at times in between. This comparison is typically assumed to be stable, that is, that one can ignore the precise $\epsilon > 0$ magnitude expressed by 'shortly before', safely assuming that if ϵ is sufficiently small, the values of the probabilities will retain the property of interest, an inequality in this case. The inequality must be assumed to hold for all values closer than ϵ because otherwise its holding would be completely arbitrary—it could be made to hold or not depending on the particular ϵ one chose, which would render the inequality meaningless for the purpose at hand: there is a very important difference between it holding for some $\epsilon > 0$ and it holding for some $\epsilon > 0$ and all smaller.

This kind of stability can be assumed in general only if the probability trajectories are continuous with respect to time to the left of t_c and/or t_e. I focus below on time t_e, but the same reasoning applies to time t_c or any other time between them. The relevant probability inequality is:

$$P_{t_e-\epsilon}(e|c) > P_{t_e-\epsilon}(e| \sim c). \tag{3.1}$$

The critical probabilities are 'shortly before' the time of the putative effect *e* because at the precise time of *e* the (conditional) probabilities are trivial. If the probability trajectories involved are not continuous to the left of t_e, then the mere fact that the inequality holds at a given time 'shortly before t_e' fails to ensure that it will hold (to the left) in any interval about t_e. If the inequality could be reversing in the neighborhood $(t_e - \epsilon, t_e)$, then it holding at t_e–ϵ is not going to be decisive for the causal efficacy of *c*, since such accounts clearly require a non-arbitrary sense of Equation (3.1) for their warrant and plausibility.

Luke Glynn's ([2011]) admirably complete account shows that even when utilizing variables instead of events for the relevant probability assessments, there remains a dependence on time, and hence continuity. Glynn originally employed a 'just before' ϵ-inequality in his 'A Probabilistic Analysis of Causation', however, the version published as (Glynn [2011]) eliminates such explicit reference to time, expressing causal conditions instead in terms of the conditional probabilities of variables attaining a value.[18] Nonetheless, a temporal index plays a role in the definition of Glynn's 'Revealer of Positive Relevance' set, which is to 'include only variables representing events occurring no later than t_E' (Glynn [2011], p. 358). Glynn also employs a condition reminiscent of Menzies and Kvart in requiring that there be the right combinations of 'increasers' (supporters of the inequality) and 'decreasers' (under-cutters of the inequality) in the interval (t_c, t_e). While this might be thought to stabilize the inequality like continuity does, the notion of the 'right combination' is only coherent if the set of potential increaser/decreaser points is a finite set, which is certainly not the case for the interval (t_c, t_e). Finally, in his discussion of the 'Hiker Ducking Boulder' example, Glynn proceeds by 'interpolating a variable' along the route of the boulder by which time it is too late for the Hiker to duck (Glynn [2011], p. 382). So indeed, even when utilizing 'variables attaining values' instead of 'events', temporal indices and their attendant continuity issues still loom.

---

[17]Christopher Hitchcock there develops a proposal growing out of Ned Hall's suggestion that one evaluate the probability of an effect shortly before the time at which the effect occurs (Hitchcock [2004], p. 414).

[18]The original version is still available online at http://web.mit.edu/gradphilconf/2008/A%20Probabilistic%20Analysis%20of%20Causation.pdf.

Continuity assumptions may also come into play in related discussions, for example, generalized causal relevance (Hitchcock [1993]), deterministic chance (Glynn [2010]), or rational belief revision (van Fraassen [1984]). In the first, Hitchcock's generalized account of causal relevance utilizes probability spaces and measures to define conditional probability functions on variables (for example, amount of medicine, blood pressure). There he analyzes (certain) causal claims as captured by conditions like: "'there exists an m such that $x > m$ implies that $f(x) > 1 - \epsilon$", where the value of ϵ is typically left vague' (Hitchcock [1993], p. 350). Glynn uses the inequality $Ch_{(t_c-\epsilon)w}(p_e|p_c) > Ch_{(t_c-\epsilon)w}(p_e| \sim p_c)$, translating it as, 'just before c occurred, the chance of e conditional upon the occurrence of c was greater than the chance of e conditional upon the non-occurrence of c' (Glynn [2010], p. 74). Finally, van Fraassen employs credence functions indexed by times toward the future, where P_t is the agent's credence function at time t, and P_{t+x} is her function at a later time $t + x$ (van Fraassen [1984], p. 244).

To summarize, if probability trajectories cannot be assumed to be continuous, then probabilistic accounts of causation are undermined because the probability around the times of interest are rendered potentially 'unstable' in the sense of jumping between values that may or may not preserve the relevant features of the probabilistic analysis—typically probability increases in the presence of the putative cause expressed in the form of an inequality.

A glaring question at this point might be: why not simply require that probability trajectories be continuous? There are two reasons to worry about this move. The first immediate reason is because it seems that some kinds of events must be understood as having discontinuous probability trajectories, for example quantum events. I take this up at length in the next section. A second perhaps less obvious reason is that such a continuity assumption would 'definitionally' legislate *a priori* against a particular qualitative feature of probability trajectories (a discontinuity) that may well turn out to be relevant to causation and other empirical and metaphysical questions. Whatever the remedy, it ought not be so restrictive as to decide such substantive empirical or philosophical questions by definition.

4   The Continuity Bind

So far if one is 'keeping score', the tally would seem to favor continuity. One important and influential analysis of causation that has understood causation in terms of discontinuity (Eells [1991]), but it was ultimately discontinuity issues that saw it falter. Add to this the very real (if neglected) continuity needs of many other probabilistic causal analyses and the balance would seem to tip toward continuity. But enter quantum theory. In what follows I consider the case for discontinuous probability trajectories based on quantum phenomena, which effectively 'ties the score' and creates a real 'continuity bind'. I then sketch two possibilities for mediating between the demands for continuity assumptions and the need not to rule out the possibility of discontinuity.

Quantum events like the decay of an atom may well have a non-trivial probability of occurring that does not change through time. Accordingly, at the instant they occur their probability trajectory will jump from a constant value to one as in Figure 4. Quantum events seem to be of a singular kind that does not depend on any 'ordinary' causal factors, and hence have a probability trajectory that does not 'evolve through time' until it jumps to 1. One might suppose that these discontinuities could be limited to the quantum level, but this is not obviously possible. A straightforward example suggesting otherwise involves nothing more than a Geiger counter that emits a clicking sound (macro-level event) when a micro-level decay event is detected.[19]

---

[19]Another now classic (if rather mean) example used in this context by Dretske and Snyder ([1972]) with a debt
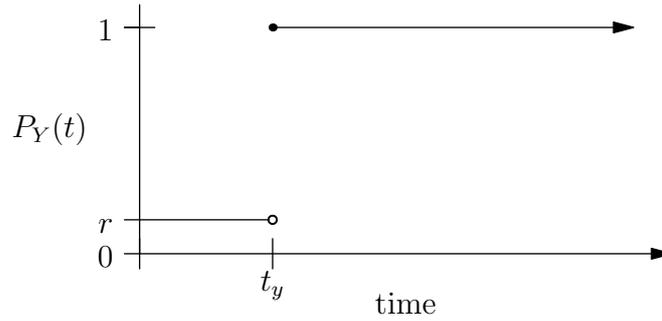
Figure 4: The discontinuous probability trajectory of a quantum-level event.

Another response might be to maintain that while at the quantum-level such quantum events have discontinuous probability trajectories, macro-level events that involve them 'dampen out' the discontinuity. On such a view events at the macro-level always have duration; they consist of intervals of time (and space). Thus the discontinuity is avoided at the macro-level because the detection event and the ensuing clicking event have temporal duration during which the probability of the click (detection event) can increase sharply but continuously to 1. Let such a quantum-level decay event have a probability of r, and then the graph in Figure 5 represents the probability trajectory of Y = the click of the counter, with interval $[t_y, t_y + \delta]$ being the duration of the detection event, that is, the time from the decay event through the detection and ensuing click. The probability of Y prior to t_y is $r - \epsilon$, for some $\epsilon > 0$, which incorporates the possibility of a mechanical or other macro-level failure to detect the particle or to bring about the click.
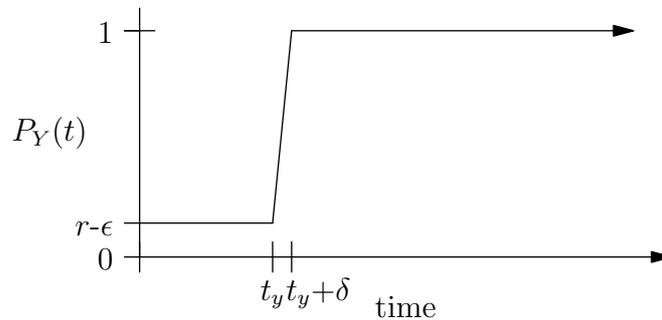

Figure 5: The continuous probability trajectory of the detection of a quantum event.

At best this response, which understands macro- and quantum-level (physical) probability and events as distinct kinds, with macro-level events having a duration that 'dampens out' the discontinuous probability at the quantum-level, saves only macro-level continuity—and with the cost of assuming a bifurcated view of physical probability and/or events. And more problematically, it involves making significant assumptions about how empirical theory will ultimately unfold. Retaining at least the possibility of discontinuous probability trajectories at all levels would seem to be the preferable way to proceed.

Thus we have the Continuity Bind. There are pressing needs for both continuity and the possibility of discontinuity. In particular:

- requiring systematic discontinuities like (DJP) or (CDJP) is problematic,

---

to Schrödinger involves a quantum mechanical process and detector hooked to a device to 'fire a revolver' at a cat, and calibrated to do so with a (quantum) probability of 0.01.

- presupposing everywhere continuous probability trajectories is too restrictive, and yet
- probabilistic causal analyses require some of the stability provided by continuity assumptions.

Next I sketch two possibilities for mediating between the demands for continuity assumptions and the possibility of (empirically or metaphysically motivated) discontinuity, namely, by (1) utilizing discontinuous first derivatives of everywhere continuous probability trajectories, or (2) abandoning point probability trajectories altogether in favor of imprecise (interval) probability trajectories.

4.1 Retaining continuity with discontinuous first derivatives

If it turns out that some events falling into place, in particular probabilistically significant ones, need in some cases to involve some kind of qualitative 'shift' in the relevant probability trajectory, then as we have seen in Eells' work, this shift can be understood as a jump discontinuity, but there is another way. An alternative is to capture such a shift by a jump in the rate of change of the probability trajectory. Formally, this could be put in terms of the first derivative (with respect to time) of the probability trajectory function, $P'_Y(t)$, which is interpretable as the rate of change of the original function $P_Y(t)$. The role played by jump discontinuities in $P_Y(t)$ in causal analyses could be handled in a parallel way by jump discontinuities in $P'_Y(t)$, and the jumps in the first derivative would still correspond to and be grounded in the idea that "causes" constitute shifts in the probability trajectory—in this case the shift would not be a discontinuous jump in the probability trajectory itself, but rather a discontinuous jump in the rate of change of the probability trajectory.[20] The probability trajectory $P_Y(t)$ could then be assumed/required to be continuous, thereby avoiding the discontinuity issue altogether.

There would still be significant work to be done developing the precise conditions, since such conditions would be based in both (1) the discontinuities of the derivative of the probability trajectory, and (2) the shape of the now continuous probability trajectory itself, and these may or may not be easy to put together. In particular, for an event to be a causal factor it would have to result in the appropriate jump in $P'_Y(t)$, but it would also have to be the case that $P_Y(t)$ itself 'remained high' as per (Eells ([1991]), Ch. 6, esp. Sec. 6.2). Another issue to work out would be precisely how differentiable $P_Y(t)$ would have to be. It would not be necessary to require $P_Y(t)$ to be everywhere differentiable; as long as the left and right derivatives existed at the point in question, the derivative itself could fail to exist.[21]

While this approach may have promise for a causal accounts, and especially for Eells' discontinuity problems, there is little reason to think that it is viable or even coherent for quantum mechanics. And further, as discussed above, a continuous probability trajectory would at best alleviate the need for macro-level discontinuities, and while the idea of entirely distinct kinds of (physical) probability at the quantum and macro-levels is not incoherent, it is not an attractive philosophical commitment as it legislates too much from a pre-empirical framework.

---

[20]There is no concern that such discontinuities in the derivative of $P_Y(t)$ might result in the same or similar problems because there is an important difference: the arguments above cannot be applied to the derivative of $P_Y(t)$ because the derivative of a probability function is not itself a probability function.

[21]Since the number of possible configurations of a probability function and its derivative afford many more complexities than are available when simply classifying discontinuities, it could well be that, as in the account of Hitchcock ([1993], p. 359), there is 'no natural division of causal relevance into a few simple species, such as positive and negative; rather, causal relevance is infinite in variety'. On this account causal relevance is infinite in variety because it is represented as an array of conditional probability functions that themselves have infinitely many possible configurations.

4.2   Imprecise (interval) probability trajectories

Within the setting of point probabilities, the pull toward and push away from continuity does indeed constitute a bind, but this is not necessarily so in settings in which probabilities are not points, but rather intervals. Imprecise (non-point-valued) probability has been studied for some time in applied and subjective probability settings, for example, (Walley [1991]), (Kyburg [1999]), and (Weichselberger [2000]), and are of renewed interest of late; see (Augustin *et al.* [2014]). And recently imprecise probabilities have been further extended to objective understandings of chance (physical probabilities); see (Glynn [forthcoming]) and (Peressini [2016]).

The advantage of interval-valued probability is that the notion of a continuous function opens up when the function in question is an interval-valued function. It turns out that there are multiple ways to generalize the standard point function definition of continuous, and thus one could seek to develop and employ a kind of continuity that

- is not so restrictive as to decide substantive philosophical questions by definition,

- stabilizes causally salient probability inequality claims between trajectories, and

- retains the possibility of 'jumpiness' to capture quantum or other theoretically motivated discontinuity.

The area in which to seek more general notions of continuity is generalized set-valued analysis. In set-valued analysis one finds weaker notions of continuity, for example, upper and lower semicontinuity, that could be employed in a way that minimizes the restrictions on discontinuities, actually allowing some kinds, and yet offering sufficient stability for causal considerations.[22]

Another approach would be to utilize classes of interval functions generated by distinct generalizations of continuity. For example, Anguelov *et al.* [2006] develops three distinct notions of continuity: S-continuity, D-continuity, and H-continuity that apply to interval functions. The class of S-continuous interval functions seem especially well-suited. And Anguelov [2004] explores how by pairing a lower semi-continuous function $\underline{f}$ with an upper semi-continuous function $\overline{f}$, such that $\underline{f} \le \overline{f}$ produces an interval function $F = [\underline{f}, \overline{f}]$, that is a completely novel entity from both algebraic and topological points of view. Such functions can be quite 'jumpy' (discontinuous in the ordinary sense) with the caveat being that the upper endpoint function can only jump up and the lower endpoint function can only jump down, and hence such functions do not have the problematic 'gaps' that discontinuous point-valued functions can have.

5   Concluding Remark

I hope to have shown that the question of the continuity of probability trajectories does indeed have important implications for any account of causality that employs a form of the 'probability raising' condition. If the probability functions can be 'jumping about' in ways typical of discontinuous functions, then the stability of the relevant probability increase is called into question. Because of this, coupled with pressure from physical (quantum) theory to allow the possibility of discontinuities, one is faced with a (dis)continuity bind that appears to be difficult to resolve in the standard framework. While a continuous trajectory with discontinuities in its

---

[22]See (Peressini [2016]) for further details.

first derivative doing the work is one possibility, exploration of the imprecise probability framework seems the more promising option for a way out of the bind, that is, a way of retaining the possibility of discrete 'shifts' in a qualitative feature of a probability trajectory while still enjoying some of the stability of continuity.

Appendix A   Proof of the Inconsistency of C/DJP

This appendix contains the detailed proof sketched in Section 2.3.1.

Let $0 < L < 1$ be the left hand limit of $P_Y(t)$ at the time Y occurs, t_y. To show that $P_Y(t)$ is discontinuous from the left at t_y it is sufficient to find a sequence of times $\{t_i\}$ converging to t_y such that $P_Y(t)$ evaluated at those points does not converge to L, that is, $\{P_Y(t_i)\} \not\to L$. In terms of the definition of convergence this means showing that there is an $\epsilon > 0$ such that

$$|P_Y(t_i) - L| \geq \epsilon, \quad \text{for some } i \text{ greater than any } N > 0. \tag{A.1}$$

The next task is to actually construct the problematic sequence $\{t_i\}$ and show that it satisfies Equation (A.1).

Since by hypothesis $\{D(X_i, \hat{t}_i)\}$ does not converge to zero, there is an $\hat{\epsilon} > 0$ such that

$$D(X_i, \hat{t}_i) \geq \hat{\epsilon}, \quad \text{for some } i \text{ greater than any } N > 0. \tag{A.2}$$

It will simplify notation to define $\{L_i\}$ to be the sequence of left hand limits for each of the $\{X_i\}$, so $L_i = 1 - D(X_i, \hat{t}_i)$, from which it follows that $|1 - L_i| \geq \hat{\epsilon}$ because of Equation (A.2). To construct the sequence of times, $\{t_i\}$ that will generate the contradiction, we must find a sequence of moments slightly before each of the $\{\hat{t}_i\}$, since we will be interested in what is happening to the probabilities of the $\{X_i\}$'s right before the time they occur—recall that $P_{X_i}(\hat{t}_i)$ is simply equal to 1. A natural choice would be to pick moments like the following: $\hat{t}_i - \frac{1}{10^i}$. Such a sequence of moments are always just before the moment $\{X_i\}$ occurs and are such that as $i \to \infty$, they get arbitrarily close to those moments, $\hat{t}_i$. But for purposes here, we need the new sequence to be more "tightly" tied to $P_{X_i}(t)$.

The factor required will depend on each of the $P_{X_i}(t)$ and how quickly each approach their limit near $\hat{t}_i$. The limit in question is L_i, so by the definition of limit, we can find a minimal distance $\delta > 0$ such that the distance between $P_{X_i}(t)$ and L_i can be made less than an arbitrary $\epsilon > 0$ for all t within the minimal distance δ of $\hat{t}_i$. The ϵ we will use here comes from the assumption that $D(X_i, \hat{t}_i) \not\to 0$, namely, $\hat{\epsilon}/2$, which is half of the $\hat{\epsilon} > 0$ we have from Equation 2.5. Now consider the sequence of $\{\delta_i\}$, each greater than zero, with the property that if $|\hat{t}_i - t| < \delta_i$, then $|L_i - P_{X_i}(t)| < \hat{\epsilon}/2$. Again, that the $\delta_i > 0$ exist follows from the definition of the (left) limit of $P_{X_i}(t)$ at $\hat{t}_i$ being L_i. Finally we define the $\{t_i\}$ as: $t_i = \hat{t}_i - \frac{\delta_i}{10^i}$. Thus the sequence of $\{t_i\}$ are such that as $i \to \infty$, the t_i's get arbitrarily close to the $\hat{t}_i$'s (the times that the X_i's occur), and further, each t_i is within δ_i of $\hat{t}_i$, so $|L_i - P_{X_i}(t_i)| < \hat{\epsilon}/2$, for each t_i. This entails that

$$|1 - P_{X_i}(t_i)| = |1 - L_i + L_i - P_{X_i}(t_i)| \geq |\,|1 - L_i| - |L_i - P_{X_i}(t_i)|\,|,$$

after adding and subtracting L_i, regrouping, and making use of the identity: $|a + b| \geq ||b| - |a||$. Because from above $|1 - L_i| \geq \hat{\epsilon}$ and $|L_i - P_{X_i}(t_i)| < \hat{\epsilon}/2$, we have that

$$|1 - P_{X_i}(t_i)| \geq |\hat{\epsilon} - \hat{\epsilon}/2| = \hat{\epsilon}/2. \tag{A.3}$$

for some i greater than any $N > 0$. This step also depends on the fact that if $|a| \geq |c|$ $|b| \leq |d|$, then $|a - b| \geq |c - d|$. Finally, using Equation A.3, removing unnecessary absolute value signs, and relabeling our crucial value $\hat{\epsilon}/2$ as ϵ for notational simplicity we get:

$$1 - P_{X_i}(t_i) \geq \epsilon, \quad \text{for some } i \text{ greater than any } N > 0. \tag{A.4}$$

In making use of Bayes' Theorem below, the conditional probabilities $P_{YX_i}(t_i)$ and $P_{X_iY}(t_i)$ will be needed. It will be assumed that $\lim_{i \to \infty} P_{YX_i}(t_i) = L$ and $\lim_{i \to \infty} P_{X_iY}(t_i) = 1$. The reasoning for this is as follows: given that the limit of $P_Y(t_i)$ is L, conditionalizing on the X_i's, which are the particular events leading up to Y, should not affect convergence; similarly since (by definition) the $\{X_i\}$ are events leading up to Y at times $\{t_i\}$ converging to the time they occur $\hat{t}_i$, the probability of these events at these times conditionalized on Y will naturally converge to 1 as time converges to the time of Y.[23] The convergence of the conditional sequence $P_{YX_i}(t_i)$ to L entails that for any $\epsilon_1 > 0$ there is an $N_1 > 0$ such that for all $i > N_1$

$$|P_{YX_i}(t_i) - L| < \epsilon_1 \tag{A.5}$$

Similarly, the convergence of the conditional sequence $P_{X_iY}(t_i)$ to 1 entails that for any $\epsilon_2 > 0$ there is an $N_2 > 0$ such that for all $i > N_2$

$$|1 - P_{X_iY}(t_i)| < \epsilon_2 \tag{A.6}$$

Now the particular values to be used for ϵ_1 and ϵ_2 are

$$\epsilon_1 = \frac{\epsilon L}{2(\epsilon + 2)} > 0 \quad \text{and} \quad \epsilon_2 = \frac{\epsilon}{4} > 0.$$

Thus from Equation (A.6) and the value for ϵ_2 and multiplying through by $L > 0$

$$|L - LP_{X_iY}(t_i)| < \epsilon_2 L = \frac{\epsilon L}{4}. \tag{A.7}$$

Then adding and subtracting L, using the triangle inequality, and Equations (A.5) and (A.7) we may derive that

$$
\begin{aligned}
|P_{YX_i}(t_i) - LP_{X_iY}(t_i)| &= |(P_{YX_i}(t_i) - L) + (L - LP_{X_iY}(t_i))| \\
&\leq |P_{YX_i}(t_i) - L| + |L - LP_{X_iY}(t_i)| \\
&< \frac{\epsilon L}{2(\epsilon + 2)} + \frac{\epsilon L}{4} \\
&= \frac{2\epsilon L + \epsilon L(\epsilon + 2)}{4(\epsilon + 2)} \\
&= \frac{\epsilon L(\epsilon + 4)}{4(\epsilon + 2)} \tag{A.8}
\end{aligned}
$$

Next using Equation (A.5) again with the value for ϵ_1 we have that

$$P_{YX_i}(t_i) > L - \frac{\epsilon L}{2(\epsilon + 2)}. \tag{A.9}$$

And now from Equation (A.4), multiplying by $P_{YX_i}(t_i)$, and using Equation (A.9), we can derive

---

[23]Again, nothing turns on these particular values for the limits of the conditional probabilities. As long as they converge to some $0 < \hat{L} \leq 1$, as they must, then the proof can proceed with simple scaler adjustments.

that

$$P_{YX_i}(t_i)(1 - P_{X_i}(t_i)) \geq (L - \frac{\epsilon L}{2(\epsilon + 2)})\epsilon$$

$$= (\frac{2L(\epsilon + 2) - \epsilon L}{2(\epsilon + 2)})\epsilon$$

$$= (\frac{4L + 2\epsilon L - \epsilon L}{2(\epsilon + 2)})\epsilon$$

$$= \epsilon L(\frac{4 + 2\epsilon - \epsilon}{2(\epsilon + 2)})$$

$$= \frac{\epsilon L(\epsilon + 4)}{2(\epsilon + 2)}. \tag{A.10}$$

Finally returning to Equation (A.1) which expresses the continuity condition:

$$|P_Y(t_i) - L| = \left| \frac{P_{X_i}(t_i)P_{YX_i}(t_i)}{P_{X_iY}(t_i)} - L \right| \qquad \text{(Bayes' Theorem (2.1))}$$

$$= \frac{1}{P_{X_iY}(t_i)} \left| P_{X_i}(t_i)P_{YX_i}(t_i) - LP_{X_iY}(t_i) \right|$$

$$\geq |P_{X_i}(t_i)P_{YX_i}(t_i) - LP_{X_iY}(t_i)| \qquad \text{(since } P_{X_iY}(t_i) \leq 1)$$

$$= |P_{X_i}(t_i)P_{YX_i}(t_i) - P_{YX_i}(t_i) + P_{YX_i}(t_i) - LP_{X_iY}(t_i)|$$

$$= |P_{YX_i}(t_i)(P_{X_i}(t_i) - 1) + (P_{YX_i}(t_i) - LP_{X_iY}(t_i))|$$

$$\geq |\,|P_{YX_i}(t_i)(1 - P_{X_i}(t_i))| - |(P_{YX_i}(t_i) - LP_{X_iY}(t_i))|\,| \qquad (|a + b| \geq ||b| - |a||)$$

$$\geq \left| \frac{\epsilon L(\epsilon + 4)}{2(\epsilon + 2)} - \frac{\epsilon L(\epsilon + 4)}{4(\epsilon + 2)} \right| \qquad \text{(from Equations (A.8) and (A.10))}[24]$$

$$= \frac{\epsilon L(\epsilon + 4)}{4(\epsilon + 2)} > 0. \tag{A.11}$$

This holds for some i greater than any $N > \max\{N_1, N_2\} > 0$. Thus $P_Y(t)$ is discontinuous from the left at t_y. $\qquad\square$

---

[24]This step also depends again on the fact that if $|a| \geq |c|$ and $|b| \leq |d|$, then $|a - b| \geq |c - d|$.

Acknowledgements

Marquette University
Department of Philosophy
Milwaukee, Wisconsin
anthony.peressini@marquette.edu

References

Anguelov, R. [2004]: 'An Introduction to Some Spaces of Interval Functions', *arXiv Mathematics e-prints*, accessed December 15, 2015 at `http://arxiv.org/abs/math/0408013`.

Anguelov, R., Markov, S. and Sendov, B. [2006]: 'The Set of Hausdorff Continuous Functions—The Largest Linear Space of Interval Functions', *Reliable Computing*, **12**(5), pp. 337–63.

Augustin, T., Coolen, F., de Cooman, G. and Troffaes, M. [2014]: *Introduction to Imprecise Probabilities*, Wiley & Sons.

Collins, J., Hall, N. and Paul, L. A. (*eds*) [2004]: *Causation and Counterfactuals*, Cambridge MA: MIT Press.

Dretske, F. I. and Snyder, A. [1972]: 'Causal Irregularity', *Philosophy of Science*, **39**(1), pp. 69–71.

Eells, E. [1991]: *Probabilistic Causality*, New York: Cambridge University Press.

Eells, E. [2010]: 'Objective Probability Theory Theory', in (Eells and Fetzer [2010]), pp. 3–44.

Eells, E. and Fetzer, J. (*eds*) [2010]: *The Place of Probability in Science*, Netherlands: Springer.

Glynn, L. [2010]: 'Deterministic Chance', *British Journal for the Philosophy of Science*, **61**(1), pp. 51–80.

Glynn, L. [2011]: 'A Probabilistic Analysis of Causation', *British Journal for the Philosophy of Science*, **62**(2), pp. 343–92.

Glynn, L. [forthcoming]: 'Unsharp Best System Chances', preprint accessed December 5, 2015 at http://philsci-archive.pitt.edu/10239/.

Hájek, A. [2003]: 'What Conditional Probability Could Not Be', *Synthese*, **137**(3), pp. 273–323.

Hall, N. [2004]: 'Two Concepts of Causation', in (Collins *et al.* [2004]), pp. 181–276.

Hitchcock, C. [1993]: 'A Generalized Probabilistic Theory of Causal Relevance', *Synthese*, **97**(3), pp. 335–64.

Hitchcock, C. [2004]: 'Do All and Only Causes Raise the Probabilities of Effects?', in (Collins *et al.* [2004]), pp. 403–17.

Ismael, J. [2011]: 'A Modest Proposal About Chance', *Journal of Philosophy*, **108**(8), pp. 416–42.

Kvart, I. [2004]: 'Causation: Probabilistic and Counterfactual Analyses', in (Collins *et al.* [2004]), pp. 359–86.

Kyburg, H. [1999]: 'Interval-Valued Probabilities', accessed online, December 15, 2015, http://www.sipta.org/documentation/interval_prob/kyburg.pdf.

Lewis, D. [1986]: 'Postscripts to 'Causation'', in D. Lewis (*ed.*), *Philosophical Papers Vol. II*, Oxford University Press, pp. 172–213.

Menzies, P. [1989]: 'Probabilistic Causation and Causal Processes: A Critique of Lewis', *Philosophy of Science*, **56**(4), pp. 642–63.

Menzies, P. [1996]: 'Probabilistic Causation and the Pre-Emption Problem', *Mind*, **105**(417), pp. 85–117.

Noordhof, P. [1999]: 'Probabilistic Causation, Preemption and Counterfactuals', *Mind*, **108**(429), pp. 95–125.

Northcott, R. [2010]: 'Natural-Born Determinists: A New Defense of Causation as Probability-Raising', *Philosophical Studies*, **150**(1), pp. 1–20.

Pearl, J. [2009]: *Causality*, Cambridge University Press.

Peressini, A. F. [2016]: 'Imprecise Probability and Chance', *Erkenntnis*. Published online September 1, 2015, http://dx.doi.org/10.1007/s10670-015-9755-9.

Rosen, D. A. [1978]: 'In Defense of a Probabilistic Theory of Causality', *Philosophy of Science*, **45**(4), pp. 604–13.

Royden, H. L. [1988]: *Real Analysis*, New York: Macmillan, second edition.

Rudin, W. [1976]: *Principles of Mathematical Analysis*, New York: McGraw-Hill, third edition.

Sober, E. [2010]: 'Evolutionary Theory and the Reality of Macro Probabilities', in (Eells and Fetzer [2010]), pp. 133–62.

van Fraassen, C. [1984]: 'Belief and the Will', *Journal of Philosophy*, **81**(5), pp. 235–56.

Walley, P. [1991]: *Statistical Reasoning with Imprecise Probabilities*, London: Chapman and Hall.

Weichselberger, K. [2000]: 'The Theory of Interval-Probability as a Unifying Concept for Uncertainty', *International Journal of Approximate Reasoning*, **24**(2–3), pp. 149–70.

Woodward, J. [2003]: *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press.