Probability from Possibility: a counterfactual definition of one kind of probability

Adam Morton

abstract: Both probability and possibility both come in degrees. This is central to the applications of probability, and the relation between possible words when one is nearer to actuality than another is implicit in counterfactual conditionals and many other modal ideas. I give a common structure of which both are special cases, more specifically, an application of the relation underlying counterfactuals within which probabilities can emerge.

I can share this draft with anyone interested. Email me at *adam.morton@ubc.ca* .

Many propositions, events, and situations can be massaged into orderings that look formally like probabilities. Strength of belief, proportions, dispositions. My aim here is to describe how one familiar use of "probability" can be connected with something equally familiar, saying what would have been the case had some event occurred. I will define a kind of probability in terms of the semantical apparatus behind counterfactual conditionals[1]. It may not be surprising that often when we speak informally of what is probable and of what would happen were some conditions met we are saying similar things. We speak of what is likely in both cases; we describe what could easily happen and what is a very remote possibility; we have idioms like "if – and it is a big if –" which indicate that a condition would only be realized under exotic circumstances. And we have idioms that mix probabilistic and counterfactual ideas: "if you had continued to bet on the wheel, your run would have ended eventually", "if you had continued to select you would have got a more typical sample", "if you had picked a person at random then they would most likely have revealed a similar prejudice". But there are problems following up the suggestion.

Two obstacles and a target A die is more likely to land with an odd numbered face upwards than with 6 upwards. It is more likely that either of these will occur than that it

1 There is an inverse project of defining modal concepts, such as causation and the counterfactual, in terms of probability. Pearl (2000) defines the counterfactual in terms of causation and probability. There is no reason why these projects should be incompatible. They may illuminate one another.

will crumble to dust on hitting the table. This outcome is possible, though; it is just less possible than the outcomes one might bet on. Degrees of possibility are built into counterfactual conditionals via the standard requirement that a condition is true when its consequent holds in the *nearest* worlds (situations, histories) where its antecedent holds. The relation of nearness (remoteness, accessibility) between worlds can be used to define a relation between propositions that holds when one is more easily realized that the other. The definition is

(1) $p <_{modal} q$ iff $((p \vee q) \,\&\, \sim(p \,\&\, q)) \;\Box\!\!\rightarrow p))$ where $\Box\!\!\rightarrow$ is the counterfactual.

"If only one of them were true it would be p." This will hold when the nearest p world that is not a q world is nearer than the nearest q world that is not a p world: p is more easily realized than q.

A fundamental obstacle makes it hard to apply this relation to typical outcomes differing in probability. It is false — or at any rate it can easily seem false — that if the die were to be cast the result would be Odd, and false that it would be 6, and also false that if it were either it would be Odd. For in terms of the remoteness that is relevant to conditionals both Odd and 6 look equally remote, though the first is much more probable. They are equally remote in causal terms because all it would take to make either of them occur would be for the die to be cast. Randomness would take over from there. But if randomness is irrelevant to the truth of counterfactuals then the processes distinguishing events probabilistically are often not going to be registered in such terms. Occurrences of events resulting from the same random process but with different probabilities are intuitively equally remote from actuality. (An extremely biased coin: "If I tossed it it *would* land heads". No: it *might* land tails, unlikely though this is. Do not confuse "if just one of them were true it would be p" with "if you were forced to choose one of them, it would be reasonable to choose p".)

This was a problem about the underlying intuition. The second obstacle is a little bit more formal. A loosely linked bundle of questions that are hard to think through. Theories of probability often measure the size of "events", which are sometimes thought of as sets of possible worlds. So the probability of an event might seem to be a way of measuring how many worlds it contains. But the idea runs into a number of problems.

The first is infinity. Most propositions, events, or items of this category, involve infinitely many possible worlds. (Think of all the tiny variations that can satisfy one verbal report.) And "more and "less" are tricky when applied to infinite sets. Probability theory postulates measures that can give sets of equal cardinality different values, so probability is not a purely counting matter.

The next is remoteness. To capture many importance modal distinctions we need not just sets of possible worlds but the ordering between them that is basic to counterfactuals Lewis 1973, Bennett (2003) chs 10-12.). But this suggests that not all worlds are equal, so even if we could count or measure them this information would have to be incorporated somehow.

Individuation of possible worlds is a very fundamental problem. Consider a world consisting simply of one object, which has a probability of 2/3 of turning left and a probability of 1/3 of turning right. In order to say that two out of three worlds involve left branching and one out of three right branching, we would have to distinguish between left worlds with identical content so that there were more of them than right worlds. And it is not clear that there is any such multiplicity of possibilities in this toy case (which is meant to capture an issue that would arise in more realistic cases.) The more attractive metaphysical rival has just one left world and one right world, plus a greater tendency to the first than the second. But what can this affinity consist in?

In response to this problem and this thought experiment you may think of the branching situation repeating many many times with the result that the object goes left roughly 2/3 of the time. Or you may think of many objects passing through the situation, so that two thirds them go left. But probabilities seem to apply in more singular situations. The object's tendencies may vary over time so that if it were in the situation again it would have a less than two thirds chance of going left. And the existence of other similar objects may change the chances. We do not want to postulate away these possibilities. And of course such strategies have their own infinity problems. The real precise probabilities can be separately or even grossly different from what happens in any finite run or averaging over any finite sample.[2]. And the infinity problems have their own possibility problems. Infinite numbers of objects and infinite repetitions of events may not be found in any *possible* world.

challenging an assumption

There are ways around these obstacles. Begin with a second look at the assumption that situations where the die lands 6 are no more remote than situations where it lands Even. It is true that we can as easily describe situations where it lands either way. But there are many not very remote situations which we cannot easily describe.

To fix intuitions consider not a die but a coin. Suppose it to be large in diameter and fairly thin. If it is tossed it will land either heads or tails (ignoring the measure zero possibility of balancing on its side). These outcomes are unlikely to be produced by quantum mechanical processes, at any rate not for several steps along the line. The main influences will be interactions between tiny variations in the way it is tossed and the erratic fluid dynamics of the air, a business that is almost impossible to predict but which we can take as deterministic. So there are features of the actual situation when the toss occurs which push it to heads or to tails, and there are features of the situation before the toss which also tended, conditional on the toss occurring, one way or the other. These are tiny variations on actual conditions, but crucial for the outcomes of interest. Call them *turbulences*.

If the coin is fair the turbulences that urge heads are similar in number and in how prone they are to occur to the turbulences that urge tails. If the coin is biased there are more and more accessible turbulences inclining the coin towards one direction than to the other. (Perhaps not many more and much more, but enough.) Or at any rate this is how it will be once the coin is in the air; before it is flipped there will be turbulences of the atmosphere and of the person’s muscular and neural states that will have set the ground for these heads-urging or tails-urging turbulences. When we stop thinking in such a "Laplacian" way we see that if the coin is biased then one of "if heads or tails then heads" or "if heads or tails then tails" is true, though we do not know which, in fact perhaps

2 This is understating it. See Hajek (2009).

cannot know which, for a particular coin tossed by a particular person on a particular occasion. Moreover the greater the bias towards one outcome or the other the more definite the conditional is, the less remote its antecedent is. These are small differences in remoteness, but physically real. And if the coin is fair — though absolute precise fairness may also turn out to be one of those measure zero outcomes, possible but infinitely improbable — then both of the conditionals are false. When they are both false because their antecedents are equally remote (near/accessible/attainable) then the coin is as likely to land one way as the other.

A basic observation that clarifies things here: though in general the nearest world where a & b is more remote than the nearest world where a, the turbulences are such minimal variations on actuality (though often leading to enormous subsequent changes) that that when conjoined with a proposition describing how a random process is set off, and which is discernible enough that we can deliberately make it true, the result is no more remote than that proposition itself. The turbulences are very near to the face.

This is the core of what we need to interpret probability in terms of possibility. To express it more carefully and as a step towards isolating the conditions where it works, consider the *primal randomizing situation*. It is an unusual kind of situation in that only a fixed and anticipated number of things can happen, and while we may suspect that there are reasons why one occurs rather than another we are unable to predict which this is. (And the unpredictability has a basis in a physical asymmetry. It is not *purely* epistemic.) It is involves a triggering action t, considered as a proposition, and a set of turbulent states A = $\{a_1, \ldots, a_n\}$. (A for the turbulent atmosphere of the randomizing mechanism.) Each a_i in the presence of t will lead to one of a set of possible outcomes O = $\{o_1, \ldots, o_m\}$.[3] The a_i are minimally but definitely remote from actuality, all except one are equally remote from actuality, and given any a_j it is true that $\sim a_j \mathbin{\Box\!\!\rightarrow} (\bigvee_{s:\, s \neq j} a_s)$ and $(t \,\&\, a_j) \mathbin{\Box\!\!\rightarrow} o_j$). So[4]

(1) $t \mathbin{\Box\!\!\rightarrow} (\bigvee_{s:\, s \neq j} a_s)$, for some $0 \leq j \leq n$.

(2) $\bigwedge_j (\sim a_j \mathbin{\Box\!\!\rightarrow} (\bigvee_{s:\, s \neq j} a_s))$

3 n, numbering the turbulences, is usually *much* greater than m, numbering the outcomes. This is because there are typically many turbulences that will lead to a particular outcome. *Perhaps* there can sometimes be infinitely many. This would complicate the notation, but I think it raises no difficulties of principle. We do not have to count anything.

4 There probably are tidier formulations. In fact, (3) may be enough all by itself.

(3) $\bigwedge_j ((t \,\&\, a_j) \;\square\!\!\rightarrow\; o_j)$

(4) $\bigwedge_j ((t \,\&\, {\sim}a_j) \;\square\!\!\rightarrow\; (\bigvee_{s:\, s \neq j} a_s))$

In such a situation some of the outcomes will be less remote than others, those that can be occasioned by the trigger in conjunction with their corresponding turbulences.[5] These are the more probable outcomes. So we have another ordering, defined, for the elements of the primal randomizing situation, by

$p >_{prob} q$ iff $(t \,\&\, (p \vee q)) \;\square\!\!\rightarrow\; p$

$p =_{prob} q$ iff ${\sim}((t \,\&\, (p \vee q)) \;\square\!\!\rightarrow\; p) \,\&\, {\sim}((t \,\&\, (p \vee q)) \;\square\!\!\rightarrow\; q)$

This gives the right result in many cases. It counts heads and tails of a fair coin as equally probable, and counts one head as less probable than six heads in a row (because it would take a particular combination of turbulences to produce HHHHHH, but any of several can produce H.) And it makes the favoured outcome of a biased coin more probable than the un-favoured one. These are qualitative rather than numerical probabilities, but these are adequate for many uses and avoid traps of overprecision. (But we *can* get numbers this way; just wait.)

embedded conditionals

Before tackling the quantitative case, consider a variation and extension of the procedure just described. We are going to toss a coin, that we assume to be fair, fifty times and compare the prospect of its landing heads exactly forty-five times ("b", a borderline outcome) with the prospect of its landing heads between thirty and forty times ("a", the anticipated outcome). Assume that when tossed it will land heads or tails; no other possibilities. We expect the latter possibility much more than the former. Though we cannot say that when or if the coin is tossed it *will* land between thirty and forty times, or even that it will land between thirty and forty rather than exactly 45 times, we can still explore what would have been the case if things were more determinate. In particular, we can consider situations where it is determinate that a or b will occur. These would involve possible influences that would make one of them be the one that will occur and

5 Using one set of events that are asymmetrically related to get a hold on another set which are the real objects of probability, is like the procedure in Strevens (2011) , developed further in Abrams (2012).

not the other. (To be clear, it is a matter not simply of ensuring that there is exactly one which occurs, but there being one such that it is ensured that it occurs.) Which prospect would it be? Surely a. Were there a situation that ensured one of them it would surely be a situation that ensured a; that would be the less remote of these two somewhat remote possibilities. This amounts to a conditional with conditionals embedded in both antecedent and consequent:

$[(((a \vee b) \,\&\, \sim(a \,\&\, b)) \mathrel{\square\!\!\rightarrow} a) \vee (((a \vee b) \,\&\, \sim(a \,\&\, b)) \mathrel{\square\!\!\rightarrow} b)] \mathrel{\square\!\!\rightarrow} [((a \vee b) \,\&\, \sim(a \,\&\, b)) \mathrel{\square\!\!\rightarrow} a]$

Because a and b are incompatible, this can be simplified to

$C_1(a,b)$ $\quad (((a \vee b) \mathrel{\square\!\!\rightarrow} a) \vee ((a \vee b) \mathrel{\square\!\!\rightarrow} b))) \mathrel{\square\!\!\rightarrow} ((a \vee b) \mathrel{\square\!\!\rightarrow} a)$

(We cannot simply require

$C_0(a,b)$ $\quad (a \vee b) \mathrel{\square\!\!\rightarrow} a$

Since the underlying process, the coin toss in the example, is random, although a is more probable than b it is not more remote. Worlds where either occur are equally remote, and the simple conditional above is false.)

$C_1(a,b)$ involves just two variables, because of repetition in four potential argument places. We will also need the full four-argument form.

$C_{1+}(a,b,c,d)$ $\quad ((a \vee b) \mathrel{\square\!\!\rightarrow} a) \vee ((c \vee d) \mathrel{\square\!\!\rightarrow} c))) \mathrel{\square\!\!\rightarrow} ((a \vee b) \mathrel{\square\!\!\rightarrow} a)$

$= (C_0(a,b) \vee C_0(c,d)) \mathrel{\square\!\!\rightarrow} C_0(a,b)$

There is a straightforward causal rationale for $C_{1+}(a,b,c,d)$. When it takes less to ensure that a will happen rather than b than it does to ensure that c will happen rather than d then the separation in likeliness between a and b is greater than that between c and d. You would have to take really elaborates precautions to guarantee that the coin landed exactly once in fifty tosses rather than exactly twice, but it would be less difficult to guarantee that it landed around forty-five times instead of exactly once.

This four termed comparison is the beginning of a series of increasingly complex comparisons that also have straightforward interpretations. A simple way to to find them is to continue both series with the recursion formulas

$C_{i+1}(a,b)$ $(C_i(a,b) \vee C_i(b,a)) \square\!\!\rightarrow C_l(a,b)$

$C_{i+1+}(a_1,\ldots,a_n)$ $(C_{i+}(a_1,\ldots,a_n) \vee C_i(a_{n+1},\ldots,a_m)) \square\!\!\rightarrow C_{i+1+}(a_1,\ldots,a_n)$ where $n=2^i$, $m=2^{i+1}$

$C_{2+}(a,b,c,d,e,f,g,h)$ for example would be $(C_{1+}(a,b,c,d) \vee C_1(e,f,g,h)) \square\!\!\rightarrow C_1(a,b,c,d)$. It would assert that the difference in assurance between that of a over b and that of c over d is greater than that of a over b and that of c over d. (Not comfortable English! But intelligible.) In general, $C_{i+1+}(a_1,\ldots,a_n)$ asserts that the difference in i^{th}-order differences between $(a_1,\ldots,a_s)$ and $(a_{s+1},\ldots,a_n)$ is greater than between $(a_t,\ldots,a_m)$ and $(a_{n+1},\ldots,a_m)$, where $s=2^{i-1}$, $t=2^i$. Or, recursively, $C_{i+1+}(a_1,\ldots,a_n)$ asserts that $C_{i+}(a_1,\ldots,a_s)$ by more than $C_{i+}(a_{s+1},\ldots,a_n)$.

These formulas give a hold on probability without invoking turbulences. But they have disadvantages. A crude disadvantage is that the physical circumstances they require are not always obtainable, and when they are they do not always lead to the intuitively right qualitative probabilities. A subtler consequence of this is that while the conjunction of a trigger and turbulence is no more remote than the trigger itself, the antecedents of these conditionals stray further and further away from actuality. It is hard to sense what might happen way out there.

This is not to say that this "multigrade" procedure (terminology of Morton 1975, idea from Morton 1997) cannot be combined with the simpler one earlier. We can postulate yet more subtle turbulences, which force the modal separations to be more and more refined. (The analogues of principles (1) – (4) above would be very hard to read, without some clever notation.) And since the a_i would be just tiny variations on the microstructure of the world as it is, increasingly refined distinctions between very easily realized states near the actual randomizing situation, they would impose an increasingly complex structure on the more distant, sometimes much more distant, situations where the random outcomes, the o_j, occur. (Of course the outcomes occur very near to actuality and in some cases in actuality also. But the structure is imposed on the possible worlds where these outcomes occur for more exotic reasons.)

quantitative, orderings

We now have two ways that qualitative probabilities are squeezed in the direction of numbers. The first is the ordering of modal remoteness of the turbulences, all of them very nearly actual. The second is the series of orderings of increasingly complex qualitative differences between items. I shall take these items to be the turbulences themselves. Then we can consider the two squeezes together. The first squeeze gives us the straightforward ordering of pairs, and then it continues for the complex refinements of the second squeeze.

To put this idea into practice begin with a well understood gamble and expand it by adding an actual outcome o_t if there is one (or one that is guaranteed to occur), and an impossible outcome o_f. These will be mapped to 1 and 0 respectively. The other outcomes will be mapped to fit between them in accordance with their ordering and so that the separations between pairs of them, pairs of parents of them, and so on fits the orderings of the rest of the series. Then add further outcomes, which can all be taken from the full output of a randomizing device, with repeated trials as outcomes. These are then mapped onto the interval satisfying the same constraints.

The result a mapping from the qualitative ordering to the real line. And it is the right mapping in a way that can be seen by beginning with a standard probability distribution to an infinite set of events and applying the mapping to the greater/less order this entails. The interval between and events assigned 0.5 and any events assigned 1 will be constrained to be the same as the interval between the 0.5 events and that assigned 0. And so for the quarter points and the eighth points and so on. Eventually all separations will be pinned down correctly[6]. (One thing this reveals is that the whole series of C ordering relations is not needed for this purpose. C_1 and C_{1+} are enough in this abstract context though in real cases more might well be needed[7].)

---

6 And uniquely, as mappings to the standard real interval. Interestingly, though, there will also be mappings to nonstandard intervals for any finite series of C relations. Nonstandard assignments are interesting for two reasons. One is to argue what matters in terms of the precision of our grasp of probabilities is not literal real values from 0 to 1 but the underlying orderings of pairs and pairs of pairs and so on. This is all we need for decision-making, for example. The other is to defend the intelligibility of modal attributions with respect to many events of measure zero. The needle *could* have pointed exactly due north, even though the probability of its doing so is zero.

7 More of the full sense of physical probability might be captured with mappings to more complex structures. In earlier drafts of this paper I used lexicographical orderings, where the degrees of probability nestle between the degrees of possibility. I still think that this expresses something

beyond deterministic random processes

I have presented the counterfactuals whose (often unstateable) antecedents are at the heart of this account in terms of their consequences for the outcomes to which probability is assigned. It may well seem that on the one hand this presupposes deterministic physical laws, and on the other hand micro states that are unlikely to be subject to such laws. I do not think this is such a big worry. Counterfactuals, and for that matter many ascriptions of causation, are consistent with indeterminism. If the electron had not encountered the atom a photon would not have been emitted (even though the energy of the photon is not a determined matter). In fact the production of the electron causes the omission of the photon. Similarly, if the electrostatic and gravitational forces on an electron were just a smidge different then when it is location was measured the measurement would have been different in a minuscule way. Microstates can be fluctuations with respect to other microstates as outcomes. In this way, we can aim to enlist even the core indeterministic element of the old quantum theory and as such at the heart of the world an indeterministic place, the Schrödinger equation[8], in our cause.

That was just a promissory note and I am not going to develop the idea. A somewhat parallel worry concerns probabilities that do not have a causal origin, for example those that have their basis in statistics. The probability of an American male being 6’1” or taller is 0.166. Of course if you hang out with basketball players you will meet far more than this proportion, and if with jockeys far fewer. So this probability assumes a random selection from the whole set of American adult males. But a random selection is like a coin toss or a spin of a roulette wheel, a process whose precise results cannot be predicted, and which is shaped by tiny causes that are practically impossible to take account of. So these probabilities also can be folded in. In fact, the distribution of heights is itself of similarly shaped but unpredictable factors. So what is presented as a matter of statistics and proportions is, digging just a tiny bit deeper, what happens when you apply one causally probabilistic process to another.

---

significant, but it did not prove necessary for present purposes.

8 (Irrelevant, really) And it produces normal distributions. So there is a reason besides the central limit theorem for thinking that these are extremely basic.

conclusions

Many parts of our thinking come in degrees that can be massaged into fitting the apparatus of probability theory, closely or loosely. There are overlaps between them but no obvious reason why one should drive the others. Among them is physical probability, rooted in relatively stable but individual tendencies to behave in one way or another and revealed in stable proportions of outcomes. It is natural to sense that physical probability is linked to ideas about what could happen more or less easily and what would happen if some conditions were met. So it is tempting to work this out using the best vocabulary we have for talking about physical modality. (For all the doubts that many of us have about this vocabulary, we do not have a better alternative at this point.) That is what I have been trying to do.

Perhaps as interesting as the project of uniting or connecting different applications of probability, and probability with different conceptual resources, is the project of articulating what is different between all these uses of related words. The procedures I have used to map degrees of probability onto degrees of possibility apply only in particular cases, and even then they can connect possibility to much richer structures than the real number line usually used to evaluate probabilities. So this effort to bring things together could also be useful in keeping them apart. These are not opposed projects: only when we have a well-defined common area can we break it down systematically into parts.

(slim) **bibliography**

Abrams, Marshall (2012) Mechanistic probability. *Synthese* **187**, 2:343-375.

Hájek, A. (2009) Fifteen Arguments Against Hypothetical Frequentism. *Erkenntniss*, 70 211–235.

Bennett, Jonathan (2003) *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.

Lewis, David (1973) *Counterfactuals*. Oxford:Blackwell

Mellor, D. H. (2005) *Probability: A Philosophical Introduction.* London: Routledge.

Morton, Adam (1975) Complex Individuals and Multigrade Relations. *Nous* **9**, 3: 309-318

Morton, Adam (1997) Hypercomparatives. *Synthese* **111**, 1: 97-114

Pearl, J. (2000) *Causality.* Cambridge: Cambridge University Press.

Strevens, Michael (2011) Probability Out Of Determinism. In *Probabilities in Physics*, Claus Beisbart and Stephan Hartmann, eds: Oxford University Press