Stochastic Independence and Causal Connection

Michael Strevens

To appear in *Erkenntnis*

Abstract

Assumptions of stochastic independence are crucial to statistical models in science. Under what circumstances is it reasonable to suppose that two events are independent? When they are not causally or logically connected, so the standard story goes. But scientific models frequently treat causally dependent events as stochastically independent, raising the question whether there are kinds of causal connection that do not undermine stochastic independence. This paper provides one piece of an answer to this question, treating the simple case of two tossed coins with and without a midair collision.

Acknowledgments

For valuable comments, thanks to Jossi Berkovitz and the audience at the IHPST.

1. The Importance of Stochastic Independence

Without pervasive stochastic independence, statistical reasoning would be of little use to us. The principle of total evidence tells us to take everything relevant into account when reasoning inductively, as a consequence of which, if we are thinking about an event e, the probability of e that matters for us is $P(e\,|\,k)$, where k is all our background knowledge. Were this probability to have different values for every possible k—for every possible set of background knowledge—we would have to compile an immense compendium of statistical knowledge in order to comply with the total evidence principle. This is true whether the probability in question is subjective, epistemic, or physical.

Fortunately, we seem to have good reason to regard most of our background knowledge as probabilistically irrelevant to many of the events about which we wish to reason statistically, or equivalently, many such events are stochastically independent of most of the background. In such cases, almost all of the background can be ignored; the probability distribution over e and its sibling events is consequently within our human cognitive grasp (Harman 1973; Pollock manuscript).

Statistical models in science illustrate this observation by imposing physical probability distributions over phenomena that take into account only a few parameters, and thus that assume implicitly that the events they describe are stochastically independent of all properties not captured by these parameters.[1] Kinetic theory's velocity distribution over gas molecules, for example, supposes that a molecule's velocity is independent of everything except temperature.[2] Statistical representations of certain kinds of experimental error

1. Science's physical probability distributions are in the first instance over event types rather than event tokens; in what follows, I use the term "event" to refer to both type events and singular events as the context requires.

2. A more careful statement of the independence assumption would exclude quantities logically related to velocity and temperature from the independence claim: kinetic theory does not, for example, assume that a molecule's velocity is independent of its momentum. But there is no need to get delayed by such matters here.

assume that errors are uncorrelated, that is, that one error is independent of any other, as do "error functions" representing "noise" in many statistical models in the social sciences. Finally, stochastic models in population genetics make a number of independence assumptions, most notably, that one instance of a gene's being replicated in the next generation is independent of any other.

When is it possible to assume that two events or states of affairs are stochastically independent? And why does independence exist, when it does? To answer these questions, asked of the physical probability distributions ascribed by scientific theories, is of great importance for the foundations of physical probability. The matter has, however, received relatively little attention in the literature.

In a notable exception to this neglect, Pollock (manuscript) derives the following principle from his theory of "nomic probability": it is always defeasibly reasonable "to expect, in the absence of contrary information, that [events] will be statistically independent of one another" (p. 11).[3] You might of course find this thesis quite acceptable even if you do not share Pollock's interpretation of probability, and many, perhaps the vast majority, of scientists seem to regard it as, at the very least, a valid heuristic. The usefulness of Pollock's principle depends, however, on how common it is to have contrary information, and ultimately, how much contrary information there is to be had: if contrary information exists for every pair of properties or events, then a well-informed scientist will be unable to take advantage of the principle at all. What facts, then, will defeat the inference endorsed by the principle?

Accepting the received wisdom, Pollock remarks that either a causal connection or a logical connection between events will undermine the assumption of independence. You might, then, formulate a more informative heuristic as

3. Pollock says "properties" rather than "events". I remind you that by "event" I sometimes mean "event type"; with this understanding, I think that Pollock would allow my reformulation.

follows:

> Expect, in the absence of contrary information, that events that are not causally or logically connected will be statistically independent of one another.

(The "in the absence of contrary information" rider remains necessary because there are defeaters for independence not made explicit in the principle, the simplest example being statistical evidence against independence.)

This paper is driven by the following problem: often we want to assume the independence of causally connected events. Most often, the connection is by way of a common causal origin: the events have among their causes numerically the same event. The existence of a common cause frequently establishes a correlation; that barometer readings and the weather are both caused in part by states of atmospheric pressure, for example, accounts for the statistical connection between barometer drops and storms. Thus, a shared etiology can undermine independence.

Events with a common cause are, nevertheless, frequently treated as independent. There are two kinds of cases worth mentioning. First, on a liberal conception of what it is to share a common cause, almost any two events with some spatiotemporal proximity have something causal in common. On a very liberal conception, everything is perhaps causally connected by the Big Bang.

Second, even if you dismiss such worries as overly sensitive to minor causal commonalities, there are cases of independence where the causal connection is anything but small. In many games, two dice are shaken in a cup and thrown onto a table. The outcomes—one for each die—are assumed to be and in fact are independent, yet during the critical randomization operation, the shaking, the dice interact as vigorously as any two mutual causal influences.

Repeated coin tosses furnish a more subtle example: that consecutive coin tosses are made by the same person in the same frame of mind surely constitutes a significant similarity in causal history. Why think that the outcome of one such toss provides no information about the outcome of the other? Their common origin cannot be discounted on the grounds of causal unimportance: the croupier who tosses the coin determines almost everything about the initial conditions that is relevant to the outcome of a toss.

Statistical models in science provide many more examples. The position and velocity of a gas molecule are treated as independent of the positions and velocities of the other molecules, even though they are determined entirely by interaction with those other molecules. Similarly, an instance of gene replication in population genetics is treated as independent of the other instances, even though replication is determined in all cases by events in a single, densely causally connected ecosystem.

These considerations together suggest that stochastic independence can and often does exist in spite of a substantial, even an entire, overlap in causal origins. Further, in science and in everyday life we take great advantage of this fact (for more on which, see Strevens (2013)). Pollock's principle and other such heuristics, valid though they might be, cannot explain why certain causal connections do not undermine independence; nor can they tell us when we are justified in assuming independence in spite of causal commonalities.

When the physical probabilities in question are irreducible, as the probabilities of quantum mechanics are often supposed to be, the answers may be spelled out in the fundamental laws of nature. But the great majority of physical probabilities in science and in everyday life—the probabilities attached to gambling setups, actuarial probabilities, the probabilities of statistical physics and evolutionary biology, the probabilities discussed by Harman and Pollock—are not irreducible. This paper will focus on independence in these reducible distributions, that is, in physical probability distributions that are presumably grounded in, among other things, facts about the causal generation of the outcomes.[4]

4. Some philosophers deny the possibility of reducible physical probability. What they cannot deny is the existence of powerful scientific models that represent deterministically produced phenomena using probability distributions that incorporate sweeping independence judgments. The success of this practice must be explained even if reducible probability

Here is a problem, then, in urgent need of philosophical attention: state the conditions under which causally connected events may be assumed to be independent, relative to the reducible physical probability distributions imposed on them by our best scientific models and theories. Or, to take a slightly different approach, in conditions that otherwise favor independence, state the kinds of causal interaction that do and do not undermine independence.

The extent of the project is vast. In this paper, I put just a small piece in place, deriving conditions for independence among the outcomes of coin tosses causally connected in two ways: first, two consecutive coin tosses made by a single croupier, and second, two simultaneous coin tosses that collide in midair. These results enlarge upon previous work in Strevens (2003) (for an overview of which, see Strevens (2005) and Strevens (2013)). The case of colliding and other tossed coins is of course of limited interest in itself, but some more general lessons about causation and independence will be apparent.

2. The Probabilistic Dynamics of Coin Tosses

Consider the following simple coin tossing setup: a coin is tossed with a variable rotational velocity, allowed to spin for a fixed time, and then stopped. Whichever face is uppermost determines the outcome of the toss, heads or tails. Suppose that the underlying dynamics are deterministic; then, the outcome of such a toss is fully determined by the physics of the coin and a single variable initial condition, the coin's spin speed, assumed always to be in the same direction. (The physics of a real coin toss is rather more complex than this (Keller 1986; Diaconis et al. 2007), but the additional complications that would arise from a more realistic model add nothing to the understanding of independence.)

Consider a function that maps any possible initial spin speed for such

is in some metaphysical sense not "real" probability.

a toss to the outcome produced. The spin speeds are real numbers and the outcome will be represented by an integer, zero for tails and one for heads. ("Edge" is ignored.) I call this function the coin's *evolution function*; a typical form for the evolution function is shown in figure 1. To help you see its form, the area under the function is shaded.

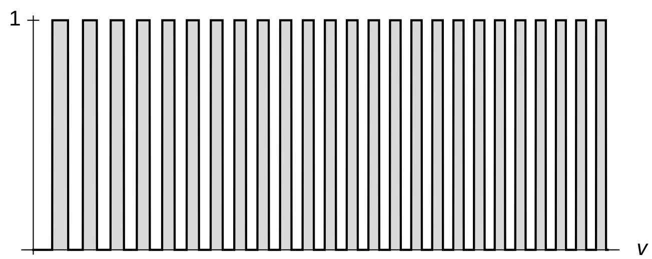

Figure 1: The evolution function for obtaining heads on a simple coin toss with initial spin speed v

The simple coin toss's evolution function has two properties that it shares with the evolution functions for many other classic gambling setups:

1. The outcome varies from tails to heads and back again very quickly as the value of the spin speed v changes, or in other words, there exist very small changes in the initial spin speed of a coin that will change the outcome of a toss from heads to tails and vice versa.

2. For any small but not-too-small interval of the spin speed v, the ratio of values of v that produce heads to those that produce tails is the same (in the case of the coin, one-to-one, that is, one-half). Or better: v can be partitioned into small intervals, in each of which the ratio of heads- to tails-producing values of v—the ratio of gray to white sections of the graph—is the same.

I call the second of these two properties, which entails the first, *microconstancy*, and I call the constant ratio of heads to tails the *strike ratio* for heads. (Formal definitions are given in Strevens (2003).)[5]

Many philosophers and scientists writing about the foundations of physical probability—von Kries, Poincaré, Hopf, and others—have thought that the microconstancy of the coin's evolution function with strike ratio one-half explains the fact that the physical probability of heads is one-half.[6] I fully agree (Strevens 2011).

For my purposes in this paper, however, it is not necessary to take a position on the metaphysics of the probability of heads. Rather, I make the following assumptions. First, in any coin toss, there is a physical probability distribution over the initial spin speed v. Second, the probability of heads is equal to the probability that the initial speed distribution assigns to heads-producing values of v.[7]

From these posits, it is possible to gain considerable insight into a problem posed above: why are the outcomes of consecutive coin tosses independent when they have a common causal origin, namely, a single human croupier?

The key to the insight is a simple mathematical theorem that also spurred the effort to ground physical probabilities in the property of microconstancy: if the evolution function of some setup is microconstant with strike ratio p for an outcome e, then any initial condition distribution, provided that it has a certain smoothness property, will induce the same probability for e, equal to p.

5. The microconstancy of an evolution function is always relative to an outcome. It is also relative to a measure, that is, a way of quantifying a setup's initial conditions. The correct measure to use in the present framework is the measure with respect to which the probability distribution over the setup's initial conditions is stated. In what follows I take the initial condition distribution, and thus the measure, as given.

6. For a guide to the older work, see von Plato (1983), and for more recent work, see Diaconis et al. (2007). Reichenbach (2008) and Hopf (1934, §9) briefly explore the relationship between microconstancy and independence, deriving a result much like that presented in the next section. For the differences between the approaches taken by the earlier writers and my own approach, see Strevens (2003, §2.A) and Strevens (2011).

7. In Strevens (2011) I endorse the second of these assumptions (in a back-handed way; see the end of §4), but I also show how to ground a probability for heads without the first assumption. In the present paper the first assumption is made for purely expository reasons.

The smoothness in question can be variously defined. One option is as follows. Microconstancy consists in the initial condition space's being divisible into small contiguous (i.e., connected) regions each with the same proportion of e-producing initial conditions. An initial condition distribution will induce a physical probability approximately equal to this proportion if it is approximately uniform—if the probability density is approximately constant—across almost all of these small regions. Call this kind of smoothness, which is sufficient but not necessary for the equality of strike ratio and probability, *macroperiodicity*.

The theorem—that any macroperiodic probability distribution over the initial conditions of a microconstant setup induces a physical probability for an outcome approximately equal to its strike ratio—is easily appreciated in pictures. Figure 2 shows two roughly macroperiodic distributions over a coin's spin speed v; in each case, the area under the distribution that is shaded gray, corresponding to the probability ascribed by the distribution to heads-producing values of v, is approximately equal to the strike ratio for heads of one-half. This is because both probability densities are (except at

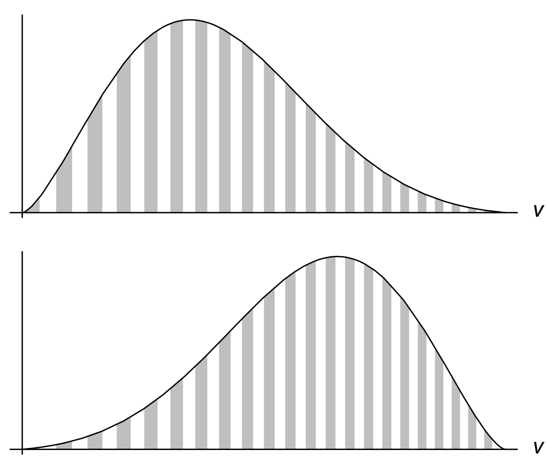

Figure 2: Different macroperiodic distributions over initial spin speed v induce the same probability for heads, equal to the strike ratio for heads

the extremes) approximately constant across each gray/white pair, so that gray and white contribute equal amounts to the total probability, yielding a probability of roughly one-half for heads and one-half for tails. The result generalizes to setups with more than one initial condition variable (and in fact can be somewhat strengthened; Strevens (2003, §2.25) states the stronger result and provides formal proofs of all results stated here).[8]

You can see from the theorem (or the figure) why two croupiers, physically different enough that they impose different physical probability distributions over the speeds with which they spin their coins, may nevertheless induce equal probabilities for heads.[9] Add a big empirical assumption to the mix—that probability distributions over spin speed tend to be macroperiodic—and you have an explanation why most tossed coins land heads about one-half of the time. These explanations are discussed further in Strevens (2003), Strevens (2011), and Strevens (2013), and also in the literature indirectly referenced in note 6. Rather than expounding further details here, however, let me turn to the question of the independence of consecutive tosses.

3. Consecutive Tosses by the Same Croupier

Consider two consecutive tosses by the same croupier. Why, despite their common causal origin, are the outcomes of the tosses independent?

You might expect a failure of independence in the following sort of sce-

8. Here and in the later results concerning probabilistic independence the term "approximate" is called on to do some hedging: approximate macroperiodicity, approximate linearity, approximate "microlinearity" (section 4), and so on. The proofs I cite and the informal arguments I give tend to assume that "approximate" means "negligible". You might wonder about cases where deviations from the ideal, though small, are not negligible. Do such deviations have a tendency to get "blown up" to big deviations by, for example, the tossed coin's sensitivity to initial conditions? They do not: the process by which the deviations are aggregated to produce the result in question (for example, a probability equal to a strike ratio) is in effect a weighted averaging; consequently, the total deviation is the weighted average of the individual deviations. That means that the total deviation cannot "blow up", and indeed will typically be much smaller than the larger individual deviations.

9. Or at least, probabilities that are almost equal.

nario. Suppose that your croupier is sometimes manic, sometimes lethargic in their coin-tossing. When manic, they tend to spin the coin faster; when lethargic they tend to spin the coin more slowly. Consecutive spin speeds produced by such a croupier will be correlated: after a fast spin, another fast spin is more likely, and vice versa. This correlation will show up in the joint probability distribution representing the spin speeds of pairs of consecutive tosses, as shown in figure 3: as you can see, there is more probability heaped around pairs of spins with similar magnitudes than pairs of spins with dissimilar magnitudes. Are the outcomes of the croupier's consecutive tosses

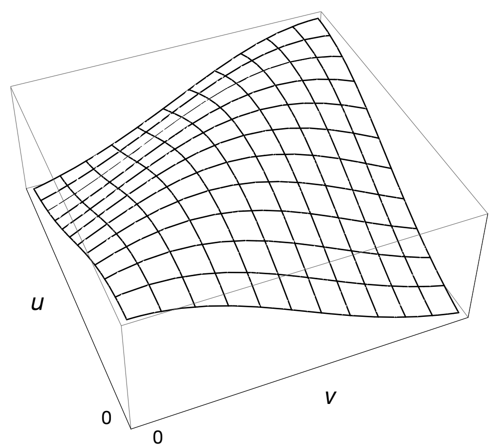

Figure 3: Joint density over the spin speeds u and v for two consecutive tosses. The magnitudes are correlated: after a fast toss, for example, another fast toss is far more likely than than a slow toss. This is represented by the "ridge" running along the $u = v$ diagonal.

correlated in the same way as the tosses' spin speeds?

They are not. Indeed, because the joint density is macroperiodic, the outcomes are not correlated at all; they are as independent as if the initial spin speeds were independent—for example, the probability of obtaining two heads in a row is, as independence requires, one-quarter. The microconstancy of the coin's evolution function renders the correlation in the tosses' initial conditions irrelevant to their outcomes.

Why? Performing two consecutive tosses can be thought of as a single "composite" process that takes as its initial conditions the spin speeds of the two tosses and produces as its outcome an ordered pair containing the outcomes of the two tosses. If the evolution function for an individual toss is microconstant with a strike ratio of one-half for heads, then the evolution function for (say) a pair of heads on the composite setup is microconstant with a strike ratio equal to the product of the strike ratios for the individual outcomes, namely, one-quarter. Such an evolution function is shown in figure 4; comparison of this graph with figure 1, or inspection of figure 5, should make it clear why the result holds.

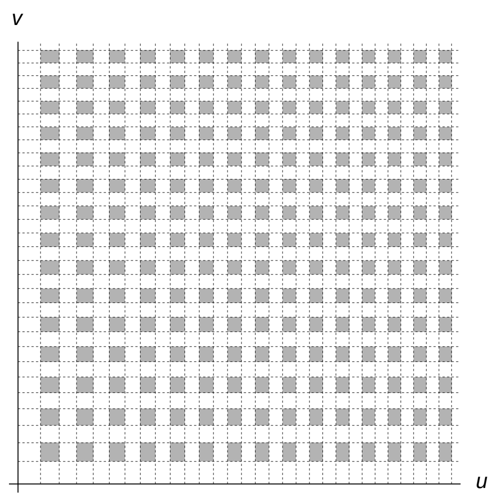

Figure 4: Composite evolution function for the event of obtaining heads on two consecutive coin tosses with spin speeds of u and v; the shaded regions show values of u and v that produce two heads

It follows that, if the joint density over the initial conditions of the two tosses is macroperiodic—like the density in figure 3—then the probability for any composite outcome is equal to the product of the probabilities for the single outcomes, as required for independence.

This result is entirely general. Given two microconstant setups, one with a strike ratio of p for outcome e and the other with a strike ratio of q for outcome

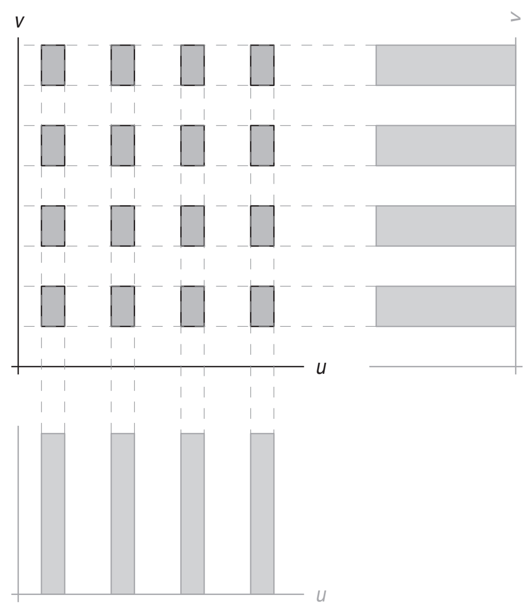

Figure 5: The relation between an evolution function for a composite event and the evolution functions for the individual events, showing that the microconstancy of the latter guarantees the microconstancy of the former with a strike ratio equal to the product of the strike ratios for the individual events. In this case the strike ratio for the event generated by initial condition u (bottom evolution function) is one-third, the strike ratio for the event generated by v (right evolution function) is one-half, and the strike ratio for the composite event (top left evolution function) is one-sixth.

f, the composite evolution function for pairs of trials on the setups will itself be microconstant with a strike ratio for the composite outcome ef of pq. (The elementary proof is given in Strevens (2003), theorem 3.3 and illustrated in figure 5.) If the joint density is macroperiodic, then the probability for the composite outcome is equal to the product of the probabilities for the single outcomes: $P(ef) = P(e)P(f)$.[10] That is necessary and sufficient for independence.

The condition required of the initial conditions by this independence result—macroperiodicity of the joint distribution—is much weaker than independence.[11] Thus, you can have a certain degree of correlation between the initial speeds of two consecutive coin tosses, induced perhaps by the croupier's state of mind or other ephemeral physiological or environmental conditions, without losing independence in the outcomes. The microconstant dynamics in effect "screens off" the correlation from the outcomes; it is an independence-creating machine, nullifying the correlating power of a common causal origin.[12]

10. Here I assume also, of course, that the densities for individual trials are macroperiodic, so that the probabilities of outcomes of individual trials are equal to the outcomes' strike ratios.

11. Mathematically speaking, almost all macroperiodic joint densities represent some degree of correlation between the events whose probabilities they encode; figure 3 depicts one example.

12. A reader asks: what if the joint initial condition distribution provides only marginal probabilities, that is (roughly) average probabilities for pairs of consecutive spin speeds? Then the independence result given above might hold but consecutive outcomes might not be independent if the single-case probabilities deviated from the averages. For example, it might be that the average probability of obtaining heads, given heads on the immediately preceding toss, is as independence requires one-half, but that in some particular case the single-case probability of heads immediately following heads is higher than one-half. (In some other particular cases it would have to be lower than one-half in order to fix the one-half value for the average.) The answer to this question is that, as with any mathematical argument, you get out what you put in. If you put in a marginal joint distribution, the argument above gets you independence of the outcomes' marginal probabilities only. If you put in a distribution of single-case probabilities, you get independence of the outcomes' single-case probabilities. If single-case probabilities exist for the outcomes, they or something equivalent presumably exist for the initial conditions (though see Strevens (2011)); the argument above

This observation goes a long way, I think, toward explaining why independence is widespread even among events that share much of their causal history. (More would have to be said, of course, to establish the existence of microconstancy in a wide range of cases of interest.)

4. Colliding Coins

The independence result for consecutive tosses stated in the previous section assumes that, although the two trials in question may have a common causal origin, they do not interact in any way once their initial spin speeds are determined. Sometimes, however, independence is found even when such interaction is present. The probabilities of population genetics, for example, represent the events over the span of a generation; within the generation, there is in most cases considerable potential for interaction between the gene-bearers. How can there be independence in spite of these causal cross-currents?[13]

In the remainder of this paper, I start on an answer to this question by examining a case in which there is a causal interaction between two tossed coins: they collide in midair. In some circumstances of practical interest, I show, the collision does not interfere with independence.

I will not be overly concerned with a realistic physics for collisions; the

should then be applied to these single-case probabilities for the pairs of consecutive spin speeds—a sufficient condition for independence in that case being the macroperiodicity of those probabilities' distribution. If single-case probabilities do not exist for the outcomes—if the only relevant probabilities are ensemble probabilities—then there can be a fact of the matter about independence only at the ensemble level; thus, it is unreasonable to request the demonstration of anything more.

13. In the treatment of tossed coins in this section, I begin with probabilities for causally isolated coin tosses—coins that do not collide—and show that most collisions do not undermine independence. That method cannot be applied directly to the probabilities of population genetics, which are inherently extrinsic, and so do not exist in a causally isolated form. The way to demonstrate the independence of the population genetic probabilities is to build them out of probabilities for smaller causal steps that do come in an isolated form, as explained in Strevens (2003, chap. 4).

aim is to survey various ways in which independence might be preserved in interacting probabilistic setups generally rather than to zoom in on the particular forms of preservation germane to spinning metal discs. I will also suppose at first for simplicity's sake that there are no frictional forces on the coins, so that they spin at a constant speed before and after any collision. (This assumption will be relaxed later.) As a consequence of this assumption, the evolution function for double heads has the perfectly regular form shown at the top of figure 6 rather than the form shown in figure 4.

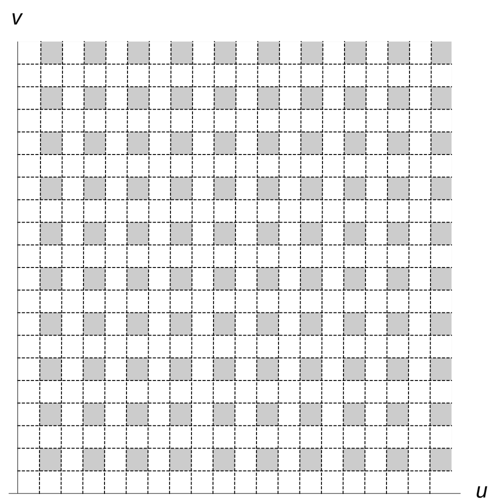

Figure 6: Evolution function for the outcome of double heads on two non-colliding coins with no friction

Two coins are tossed simultaneously. Partway through their flight they collide. What next? Perhaps they stop each other's spinning dead, and fall to the ground without any further revolutions. The effect of the collision, then, is simply to reduce the interval of time for which the coins are allowed to spin before the outcomes of the tosses are read off from their positions. If they collide halfway through their allotted spin time, for example, the effect is the same as if the allotted spin time were half of what it actually is.

What you have, then, is a case that is formally identical to the case of non-interacting consecutive tosses analyzed in the previous section. You might think that no more needs to be said: the outcomes of consecutive tosses that do not interact are independent, so the outcomes of tosses that collide in the way just described are also independent.

That would be a little too quick, however. Because the collision reduces the time for which the coins spin, it increases the width of the gray and white stripes in the relevant evolution functions; you can see the effect by comparing the evolution function for double heads without the collision (figure 6) to the evolution function for double heads with the collision (figure 7). If independence is to hold in spite of the collision, the initial spin speed distribution must be macroperiodic relative to this new evolution function as well as to the old, which means that it needs to be approximately uniform over larger regions than it needed to be to ensure the independence of consecutive non-colliding tosses. In this particular case, it needs to be uniform over regions that are something like four times the size of the regions in the non-colliding case.

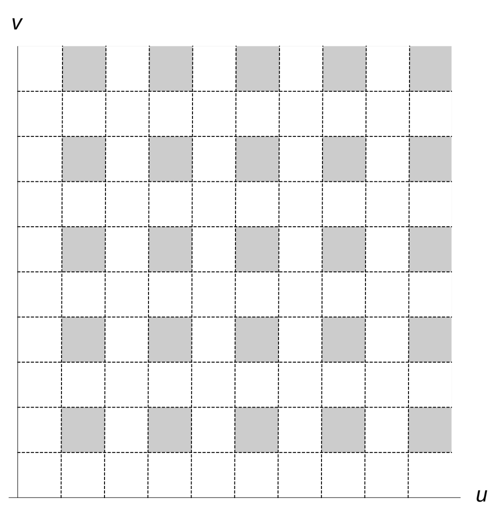

Figure 7: Evolution function for the outcome of double heads on two coins that collide in midair stopping each other dead

Let me develop a terminology for talking about this issue succinctly. Say that the *strength* of an evolution function's microconstancy is inversely proportional to the size of the small areas of constant gray-to-white ratio that make it microconstant. The evolution function for the non-colliding coins is considerably stronger in its microconstancy, then, than the evolution function for the colliding coins, because the former function can be divided into regions of constant gray-to-white ratio that are considerably smaller than the corresponding regions for the latter function.[14]

Likewise, say that the strength of an initial condition distribution's macroperiodicity is proportional to the size of the areas over which it is approximately uniform. The "smoother" or "flatter" it is in the small, then, the stronger its macroperiodicity.

To return to the colliding coins, then: the effect of a collision that stops the coins spinning is to weaken the microconstancy of (that is, diminish the strength of the microconstancy of) the evolution functions that determine the probability of double heads and so on. Such a collision will fail to undermine the independence of the tosses if the corresponding initial condition distribution is strong enough to take up the slack. But if it is not very strong, then the collision will expose some of the high-level correlation between the coins' initial spin speeds (if such correlation exists), so undermining independence.

The earlier in the toss the coins collide, the more the microconstancy of the corresponding evolution functions is weakened. Thus the following loose rule of thumb can be promulgated: a collision that brings the colliding coins spinning to a dead stop is more likely to undermine independence the earlier it occurs; accordingly, a relatively "late" collision of this sort will in many cases not interfere with independence. This offers a first example, then, of causal interactions that preserve independence.

14. This informal definition is imprecise: if the areas of constant ratio are of different sizes, for example, it does not specify whether strength is determined by taking the largest size, the mean size, or something else. But given the informal use to which I put the notion in what follows, there is no real benefit to precisifying—though a fuller treatment of independence would certainly do so.

Useful though this conclusion will be in what follows, the case itself is of limited intrinsic interest, as the causal interaction in question—the spin-stopping collision—does not induce any sort of correlation between the states of the coins. Let me next show you a case of a collision that does create such a correlation and that as a result completely undermines the independence of the outcomes of the tossed coins.

Suppose that the colliding coins fly toward one another spinning in opposite directions and that their leading edges collide with the same perfectly elastic dynamics as an idealized collision between billiard balls, so that they simply swap spin speeds. (Again, the realism of the physics is not so important here.) Suppose that this collision occurs halfway through the toss. In that case, each coin will spin at its original speed for half of the allotted time, and then at the other coin's original speed and in the opposite direction for the other half of the allotted time. If the initial speeds of the two coins are u and v, then, their mean speeds over the course of the toss are respectively $u/2 - v/2$ and $v/2 - u/2$. The magnitudes of the two mean speeds are, consequently, identical, which means that as a result of the collision the coins make exactly the same number of revolutions before landing. Their outcomes will be exactly correlated, then: if you know one, you will be able to predict the other with certainty.

In the remainder of this paper, I ask under what circumstances collisions of this sort—collisions in which the mean speed of each colliding coin is a non-trivial function of both coins' initial speeds—destroy independence. It turns out that in a wide range of cases, independence is preserved. The present case is, then, far from representative of the effect of causal commerce on independence.

Consider next a case in which the coins collide as above halfway through the allotted time but in which they slow one another down rather than entirely exchanging velocities. Suppose, in particular (and without worrying about getting the physics exactly right), that as a result of the collision the new speeds of the two coins are determined by their initial speeds u and v as follows:

$$u' = u - v/2; \quad v' = v - u/2 \tag{1}$$

The resulting evolution function for double heads is shown in figure 8. You

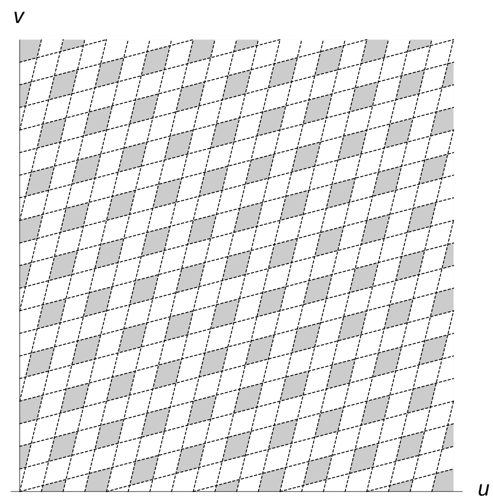

Figure 8: Evolution function for double heads with a linear collision halfway through the toss

can see right away that this new function is microconstant with a strike ratio of one-quarter, as required for independence. It seems, then, that the collision preserves independence, even as it mixes the spin speeds.

Why does the collision not undermine independence? More exactly, why does it preserve the microconstancy of the evolution function? Think of the collision as transforming the old, collision-free evolution function into a new evolution function that reflects the dynamics of the collision. In this case, the evolution function in figure 6 is transformed into the evolution function in figure 8. Now consider any small region in the transformed evolution function. I will show that, in the right conditions, the ratio of gray to white in such a region is equal to the strike ratio for the old function. Thus the new function is microconstant with the same strike ratio as the old function, as independence requires.

The small region in the new evolution function is a transformation of a region in the old function—the region that is its "inverse image". If the old evolution function is microconstant and the inverse image is not too small, then the ratio of gray to white in the inverse image is equal to the old function's strike ratio. Because the collision is linear (and invertible), the new evolution function is a linear function of the old function, and in particular the small region we are looking at is a linear transformation of its inverse image. Linear functions preserve proportions (if invertible), so the proportion of gray in the small region in the new evolution function is equal to the proportion of gray in its inverse image, thus to the strike ratio of the old evolution function, as desired.

The preservation of independence by the collision above is explained, then, by the collision's linearity. Also necessary for the explanation is, as assumed in the previous paragraph, that a small region's inverse image is not *too* small (or else its gray-to-white ratio might not reflect the old evolution function's strike ratio). What we want, roughly, is that the size of the inverse image is no smaller, or at least not much smaller, than the size of the region onto which it maps. This means that the transformation is not "inflationary": it does not inflate small regions in the old evolution function into much larger regions in the new evolution function.

There is some room for maneuver. A mildly inflationary transformation will yield a new evolution function that is microconstant, but more weakly than the old evolution function, in the sense of weakness defined above: the regions of constant gray-to-white ratio will be larger in the new evolution function than in the old. As you saw in the case of the collision that stopped the coins dead, this is not disastrous for independence provided that the initial condition distribution has strong enough macroperiodicity to take up the slack. If we have reasonably strong macroperiodicity, then, we can allow a certain degree of inflation. (In fact, the collision represented by equation 1 and depicted in figure 8 is mildly inflationary: a region in the old evolution function is mapped to a region in the new evolution function that is 1/15th larger. You can see for yourself that this is not a big deal.)

The outcomes of colliding coin tosses will be independent, then, if the collision between the coins effects a transformation of evolution functions that is (invertibly) linear and that is not too inflationary. (The "spin exchange" transformation considered above that effected an exact correlation between the coins is linear but non-invertible.)

This generalization about independence can be strengthened to apply to a wide range of non-linear transformations, namely, the transformations that I call *microlinear*. A transformation is microlinear if its effect on any small region can be approximated by a linear transformation. A one-to-one microlinear transformation will therefore preserve gray-to-white ratios, and so the informal argument for the microconstancy of the new evolution function given above will go through for microlinear transformations, provided that they are not too inflationary.[15] By way of illustration, figure 9 shows an example of the (approximately) microconstancy-preserving effect of a nonlinear but microlinear collision.

You might ask: Surely the collision in most cases establishes a correlation between the colliding coins' spin speeds? Where does that correlation go? Why does it not undermine the independence of the tosses' outcomes?[16] Here is the answer. A collision between two coins does typically correlate the coins' spin speeds. But often, it establishes only the sort of high-level correlation shown in figure 3. It correlates the speeds, then, but only in a way that the evolution functions' microconstancy renders probabilistically irrelevant, for

15. That the transformation of evolution functions is one-to-one is guaranteed by linearity but not by microlinearity. However, because the transformation in question is a one-to-one function of the inverse of the transformation that maps pre-collision speeds onto post-collision speeds, we can be sure that the sort of evolution function transformations we are talking about are one-to-one: if they were not, the spin speed transformation would not be well-defined, or perhaps I should say, would not be deterministic.

16. Thanks to Jossi Berkovitz for raising this question. The loose generalization stated in this paragraph has its foundation in the results proved in Strevens (2003), §3.B7.

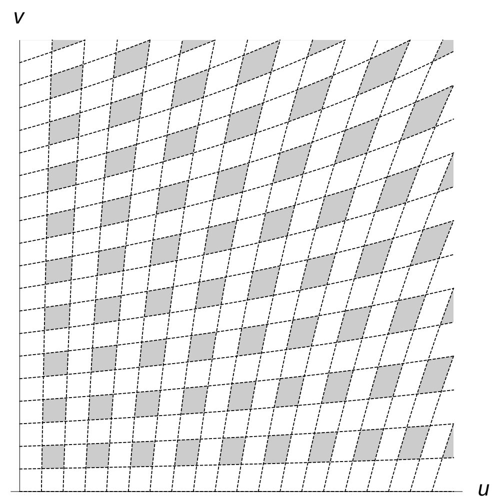

Figure 9: An evolution function for double heads transformed by a non-linear but microlinear collision. Microconstancy is (approximately) preserved.

the reasons given in section 3, to the outcomes. More specifically, it can be shown that, if the joint distribution of two colliding coins' initial spin speeds before a collision is macroperiodic, and if the collision's dynamics are of the sort specified in the preceding paragraphs—if they effect a transformation of the evolution function that is microlinear and not too inflationary—then the joint density incorporating the effect of the collision is also macroperiodic. Thus the kind of correlation effected by a microlinear, not-too-deflationary collision is the kind of correlation that, given a microconstant dynamics, makes no difference to the probabilities of the outcomes, unconditional or conditional, and so has no impact on probabilistic independence.

* * *

It would be good to have a formal result encompassing the claims made in this section. Further, it would be good to have a formal result couched, not in terms of the transformation of the evolution functions effected by the collision, but in terms of the transformation of spin speeds effected by the collision, so that we can look at a collision dynamics such as that represented by equation 1 and directly judge whether or not it undermines independence.

Such a result may be found in Strevens (2003), §3.6 and §3.B6; I state it here without proof:

> The independence of outcomes produced by simultaneous microconstant trials (with macroperiodic individual and joint initial condition distributions) is not undermined by a causal interaction, provided that (a) the interaction occurs at the beginning of the trial, and (b) the interaction can be represented by an instantaneous, non-deflationary, microlinear transformation of the initial conditions.

You will understand the significance of microlinearity in the light of the discussion above. You might wonder about the requirement that the interaction be non-deflationary, when what was required in the informal discussion above was that the transformation of evolution functions effected by the interaction be non-inflationary (where deflation and inflation are contraries). The reason for this inversion is that the transformation of old to new spin speeds (more generally, initial conditions) stands to the transformation of old to new evolution functions as a function to its inverse. One will tend to be deflationary if the other is inflationary (a tendency that will be explored more rigorously below).

The theorem is limited in two ways. First, it does not allow for mild deflation (hence mild inflation in the evolution function transformation). But you will already see how this limitation can be relaxed: a little bit of deflation in the spin speed transformation can be allowed provided that the macroperiodicity of the initial spin speed distribution is strong enough to compensate.

The more disappointing limitation of the theorem is that it applies only to interactions that occur at the very beginning of a trial, thus only to those tossed coins that collide at the first instant they take flight.[17] In the next section, I expand the result to collisions occurring at any point in the trial, so deriving a rather general set of conditions sufficient for independence preservation in colliding tossed coins and other similar systems.

5. Mid-Toss Collisions

The theorem stated above guarantees that the independence of simultaneously tossed coins will not be undermined by a collision if that collision effects a transformation of the initial conditions that is non-deflationary, microlinear, and that occurs at the first instant after the coins are tossed. My aim in this section is to remove the last restriction, showing that the result holds for a wide range of interactions at any point in the trial.

Since I am now moving into a more formal mode, some more formal definitions: A transformation is *non-deflationary* if it transforms any region into a region of equal or greater size. A simple example is $y = 2x$, which doubles the size of any interval. A transformation is *microlinear* if it can be approximated by a mosaic of linear transformations, or slightly more strongly, if for any small, contiguous region of the transformation's domain, there exists a linear transformation that has an approximately equivalent effect on the region (mapping the same points to roughly the same points). Microlinearity is of course relative to some standard for what is "micro", that is, for what counts as a small region. In the theorem, the standard is set by the relevant evolution functions' microconstancy: loosely speaking, a region is small if it is the same size, or smaller, than the regions of constant proportion in virtue of which the evolution functions are microconstant.[18]

17. For technical reasons, this created no problems in Strevens (2003).

18. More technically, the microlinearity of a transformation can be defined as relative to a partition of its domain into connected sets: it is microlinear if it is approximately linear over any member of the partition. What the theorem above requires is microlinearity relative to a partition into sets of equal strike ratio.

The structure of this section's argument is as follows: I show that under certain conditions a microlinear, non-deflationary transformation of a coin's spin speed partway through a toss is equivalent to a different transformation on its initial spin speed that is also microlinear and non-deflationary. Since the latter sort of transformation preserves independence, so does the former.

Suppose, then, that two coins are tossed, with spin speeds of u and v respectively, in such a way that they both spin for time T. At some time t_1 less than T, the coins collide, effecting a change in their spin speeds (and no other relevant physical changes). As above, I assume that a coin's spin speed does not significantly decrease over the course of the toss. (This is a significant assumption; it will eventually be dropped.)

Let me focus on the dynamics of the first coin. Supposing that spin speeds are represented in revolutions per time unit, the number of revolutions made by the first coin at the time of the collision t_1 is $t_1 u$. If the speed of the coin is altered at t_1 to u', then the number of revolutions the coin makes between t_1 and T is $t_2 u'$, where $t_2 = T - t_1$. Thus the total number of revolutions n made by the coin, which determines the outcome of this toss, is $t_2 u' + t_1 u$.

I will first consider the case where the transformation effected by the interaction is linear, and so can be represented as follows:

$$u' = au + bv + c$$

for some a, b, and c.[19] The total number of revolutions made by the coin is the sum of the revolutions made before the collision and the revolutions made after the collision:

$$\begin{aligned} n &= t_2 u' + t_1 u \\ &= t_2(au + bv + c) + t_1 u \\ &= (at_2 + t_1)u + (bt_2)v + ct_2. \end{aligned}$$

19. A non-zero value for c would be physically peculiar, but since it is easy to handle, I do not rule it out here.

My aim is to find a linear or linear plus constant transformation of u and v that yields the same formula for n when applied to u at the very beginning of the toss. This new transformation, if it exists, has the form:

$$du + ev + f.$$

If a transformation of this form is applied at the beginning of a toss, the total number of revolutions made by the coin is:

$$\begin{aligned} n &= T(du + ev + f) \\ &= dTu + eTv + fT. \end{aligned}$$

Such a transformation is equivalent to the actual transformation if there are values of d, e, and f for which these two formulas for n are equivalent. Clearly, there are:

$$\begin{aligned} d &= (at_2 + t_1)/T \\ e &= bt_2/T \\ f &= ct_2/T. \end{aligned}$$

So the effect of a linear interaction on u partway through a coin toss is equivalent to the effect of a linear interaction—though a different one—at the beginning of the toss. The same is true for v. Thus the transformation of u and v partway through the toss is equivalent to a linear transformation of u and v at the beginning of the toss.

Under what conditions will this equivalent interaction at the beginning of the toss be non-deflationary? The "Further Proofs" section at the end of this paper shows that, provided certain physically plausible conditions hold, if the transformation effected by the actual interaction of the coins is non-deflationary, then the equivalent beginning-of-the-toss transformation is also non-deflationary. It is also shown that if the transformation is deflationary, then the equivalent beginning-of-the-toss transformation is no more deflationary.

Now consider the case in which the transformation effected by the interaction is microlinear, that is, linear over any small, contiguous region of its domain. From the reasoning above, over any such region, the effect of the transformation is equivalent to a different linear interaction at the beginning of the trial applied to the same region.

It follows that the effect of a microlinear interaction partway through a coin toss, which is made up of different linear interactions over different small regions, is approximately equivalent to the effect of a corresponding set of linear interactions over the same regions at the beginning of a coin toss, thus is equivalent to a microlinear interaction at the beginning of the toss. Further, provided that the mid-toss interaction is non-deflationary, the equivalent beginning-of-toss interaction is also non-deflationary. Such an interaction will therefore preserve independence.

If the mid-toss interaction is deflationary, the equivalent beginning-of-toss interaction is no more deflationary; thus, if the mid-toss interaction's deflation is mild and the initial condition distribution's macroperiodicity is strong, independence will be preserved in this case too.

The argument as presented so far makes the rather narrow (if in many cases realistic) assumption that a coin's spin speed remains effectively constant, except where it is affected by collisions, from the beginning to the end of the toss. The assumption may, however, be relaxed, as I will now explain.

The principal part of the derivation above is the demonstration that a linear transformation of spin speeds partway through a toss is equivalent to a linear transformation at the beginning of the toss. Inspection of the proof shows that what matters for the derivation is that the formula for the number of revolutions made by the coin over a specified period of time, given an initial speed of u, is a linear function of u. This is true provided that the spin speed of the coin at any time t, given an initial speed of u, can be written $f(t)u$ (where $f(t)$ need not itself be linear), since the number of revolutions made over a period of length t is then $F(t)u$, where $F(t) = \int_0^t f(x)\,dx$. To generalize the proof above, then, simply substitute $F(t_1)$ for t_1, $F(t_2)$ for t_2, and $F(T)$ for T. The linearity condition is satisfied by a number of possible dynamics for spin speed, such as exponential decay (on which the speed at time t, given an initial speed u, is ue^{-kt}).

Two other issues must be addressed to secure independence in the case where the coin's spin speed changes with time. First, in order to have microlinearity of the beginning-of-the-toss transformation, it is not enough that the actual interaction transformation be microlinear, that is, linear over any small region of the speed space. It must be linear over the *inverse image* of any such region with respect to the speed evolution function (the function that determines how speed changes with time). A sufficient condition for this, given the microlinearity of the actual interaction transformation, is that the inverse image of any region of the speed space be at least as large as that region, because then we are guaranteed linearity over these "at least as large" sets, hence microlinearity. This sufficient condition amounts to the requirement that the speed evolution function be non-inflationary. A sufficient condition in turn for non-inflation is that $f(t)$ be a non-increasing function—an eminently reasonable supposition for a tossed coin.

The second condition for independence is that the beginning-of-the-toss transformation should be non-deflationary (or that if the transformation is mildly deflationary, the initial condition distribution's macroperiodicity should be strong enough to compensate). The "Further Proofs" show that $f(t)$'s being non-increasing is sufficient for the satisfaction of these conditions (given the other conditions already in place).

In short, then, the independence result can be extended to any case in which the speed evolution function is linear in the speed and non-inflationary. You can likely see the prospect for further generalization to speed evolution functions that are microlinear, but let me leave that to another time.

6. Conclusions

Thanks to microconstancy, there is ample independence to be found among the outcomes of coin tosses. First, microconstancy renders irrelevant many correlations due to common causal origins, as when the initial spins of consecutive tosses are correlated due to the passing moods of the croupier. Second, microconstancy protects independence against a range of causal interactions between the coins once tossed, in particular those that are non-deflationary (or not too deflationary) and microlinear.[20]

The interest of the results for coins depends, of course, on the extent of microconstancy, and in particular, on the question whether microconstancy is to be found in the kinds of systems concerning which statistical models in science make broad assumptions of independence.

I have argued for the prevalence of microconstancy elsewhere (Strevens 2003). Even so, the results referenced and developed in this paper at best constitute only the beginnings of an extensive and important philosophical project, of understanding the reasons for physical probabilistic independence among causally connected outcomes. Some further thoughts about causation and independence can be found in the material supplementary to my book *Bigger than Chaos* posted at www.strevens.org/chaos/ under "Additions". Those programmatic remarks show just how much work remains to be done.

Further Proofs: Sufficient Conditions for Non-Deflation

I will show that if a midair collision between two tossed coins effects a non-deflationary linear transformation of the spin speeds, then the equivalent transformation at the beginning of the tosses will also be non-deflationary, provided that some additional, physically plausible conditions hold. I will

20. Two additional physically plausible assumptions are made in the course of the derivation (see the "Further Proofs"): a and g are positive and b and h have the same sign.

also show that under the same conditions, a deflationary mid-toss collision is equivalent to a beginning-of-toss transformation that is no more deflationary.

Some new notation will simplify the algebra. Define $\hat{t}_1$ and $\hat{t}_2$ as the proportions of the total spin time that elapse before and after the collision, so that

$$\hat{t}_1 = t_1/T \quad \text{and} \quad \hat{t}_2 = t_2/T.$$

Then the coefficients d and e of the beginning-of-the-toss transformation of u equivalent to the actual interaction can be written

$$d = a\hat{t}_2 + \hat{t}_1$$
$$e = b\hat{t}_2.$$

The same notation can be used for the coefficients of the beginning-of-the-toss transformation of v. If the actual linear transformation of v effected by the collision is:

$$v' = gv + hu + i$$

then the beginning-of-toss transformation of v equivalent to the actual interaction is

$$(g\hat{t}_2 + \hat{t}_1)v + h\hat{t}_2 u + i\hat{t}_2.$$

Consider non-deflationary collision dynamics first. The beginning-of-the-toss transformation—the transformation on the $u \times v$ space—is non-deflationary just in case the absolute value of its determinant is greater than or equal to 1. Let me simplify matters by assuming that the determinant is positive; non-deflation in that case requires that the determinant is greater than or equal to 1 (otherwise, run the following arguments with strategically placed negation operators).

The beginning-of-toss transformation's determinant is

$$\det\begin{pmatrix} a\hat{t}_2 + \hat{t}_1 & b\hat{t}_2 \\ h\hat{t}_2 & g\hat{t}_2 + \hat{t}_1 \end{pmatrix} = ag(\hat{t}_2)^2 + a\hat{t}_1\hat{t}_2 + g\hat{t}_1\hat{t}_2 + (\hat{t}_1)^2 - bh(\hat{t}_2)^2$$

$$= (ag - bh)(\hat{t}_2)^2 + \frac{(a+g)}{2} 2\hat{t}_1\hat{t}_2 + (\hat{t}_1)^2.$$

Since $(\hat{t}_2)^2 + 2\hat{t}_1\hat{t}_2 + (\hat{t}_1)^2 = (\hat{t}_1 + \hat{t}_2)^2 = 1$, a sufficient condition for the determinant to be greater than or equal to 1 is that the coefficients of $(\hat{t}_2)^2$ and $2\hat{t}_1\hat{t}_2$ in the above expression be greater than or equal to 1, that is, that

1. $ag - bh \geq 1$, and
2. $a + g \geq 2$.

The quantity $ag - bh$ is the determinant of the actual transformation, that is, the transformation actually effected by the midair collision partway through the toss; it is greater than or equal to 1 just in case the original transformation is non-deflationary.

If it is assumed that a and g are positive (i.e., a coin's pre-interaction spin speed makes a positive rather than a negative contribution to its post-interaction spin speed) and that (because of physical symmetry) the signs of b and h are the same, then (1) entails that $ag \geq 1$, which in turn entails (2). Under these assumptions, then, if the interaction transformation partway through the toss is non-deflationary, its equivalent at the beginning of the toss is also non-deflationary.

What if the mid-toss collision transformation is deflationary? Its degree of deflation is proportional to its determinant d, the absolute value of which will be less than one. Assume as before for expository purposes that d is positive. Then it can be shown that under the same conditions assumed in the previous paragraph, the determinant of the equivalent beginning-of-toss transformation is no less than d, so that the equivalent beginning-of-toss transformation is no more deflationary than the mid-toss transformation. Proof: A sufficient condition for the beginning-of-toss determinant to be greater than or equal to d can be read off the argument above (using all variable names in the same way):

1. $ag - bh \geq d$, and

2. $a + g \geq 2d$.

Since d is by definition equal to $ag - bh$, condition (1) is always satisfied. Supposing, as before, that a and g are positive and that b and h have the same sign so that bh is positive, condition (2) is also satisfied (reasoning omitted).

Next consider the case where the coin's spin speed changes significantly over time. As in the main text, I assume that the speed at any time is a linear function of the initial speed: there exists a function $f(t)$ (not necessarily linear) such that, after time t, the speed of a coin with initial speed u will be $f(t)u$. Redefine $\hat{t}_1$ and $\hat{t}_2$ so that

$$\hat{t}_1 = F(t_1)\big/F(T) \quad \text{and} \quad \hat{t}_2 = F(t_2)\big/F(T).$$

where $F(t) = \int_0^t f(x)\,dx$. Then the relevant beginning-of-the-toss transformation can be written just as I have written it above.

It can no longer be assumed that $\hat{t}_1 + \hat{t}_2 = 1$. However, if as I suggested in the main text $f(t)$ is a non-increasing function then as demonstrated below, $\hat{t}_1 + \hat{t}_2 \geq 1$, which is sufficient for the above reasoning to apply to the more general case, so that if the other conditions stated above apply, then (a) if the actual collision transformation is non-deflationary, the equivalent beginning-of-the-toss transformation is also non-deflationary, and (b) if the actual collision transformation is deflationary, the equivalent beginning-of-the-toss transformation is no more deflationary.

To show that $\hat{t}_1 + \hat{t}_2 \geq 1$: because $f(t)$ is non-increasing, for $b > 0$

$$\int_0^a f(t)\,dt \geq \int_b^{b+a} f(t)\,dt.$$

It follows from the definition of $F(t)$ that

$$\begin{aligned}\hat{t}_1 + \hat{t}_2 &= \frac{F(t_1) + F(t_2)}{F(T)} \\ &= \frac{\int_0^{t_1} f(t)\,dt + \int_0^{t_2} f(t)\,dt}{\int_0^T f(t)\,dt}\end{aligned}$$

$$\geq \frac{\int_0^{t_1} f(t)\,dt + \int_{t_1}^{T} f(t)\,dt}{\int_0^{T} f(t)\,dt}$$

$$\geq 1.$$

References

Diaconis, P., S. Holmes, and R. Montgomery. (2007). Dynamical bias in the coin toss. *SIAM Review* 49:211–235.

Harman, G. (1973). *Thought*. Princeton University Press, Princeton, NJ.

Hopf, E. (1934). On causality, statistics and probability. *Journal of Mathematics and Physics* 13:51–102.

Keller, J. (1986). The probability of heads. *American Mathematical Monthly* 93:191–197.

von Plato, J. (1983). The method of arbitrary functions. *British Journal for the Philosophy of Science* 34:37–47.

Pollock, J. (Manuscript). Probable probabilities. Available at http://philsci-archive.pitt.edu/3340/1/Probable_Probabilities.pdf.

Reichenbach, H. (2008). *The Concept of Probability in the Mathematical Representation of Reality*. Translated and edited by F. Eberhardt and C. Glymour. Open Court, Chicago. Reichenbach's doctoral dissertation, originally published in 1916.

Strevens, M. (2003). *Bigger than Chaos: Understanding Complexity through Probability*. Harvard University Press, Cambridge, MA.

———. (2005). How are the sciences of complex systems possible? *Philosophy of Science* 72:531–556.

———. (2011). Probability out of determinism. In C. Beisbart and S. Hartmann (eds.), *Probabilities In Physics*. Oxford University Press, Oxford.

———. (2013). *Tychomancy: Inferring Probability from Causal Structure*. Harvard University Press, Cambridge, MA.