3.7A Non-Microconstant Outcome Dependent IC-Evolution

Sufficient conditions for the stochastic independence of chained non-microconstant trials

Version of 5/15/03

The Problem

In section 3.76 of *Bigger than Chaos*, I state sufficient conditions for chained trials on the same microconstant experiment each to have the same probability distribution over their outcomes, regardless of previous outcomes. This is a sufficient condition for the independence of the trials.[1] In this section I take the same approach to non-microconstant experiments, that is, I will look for sufficient conditions for chained trials on the same non-microconstant experiment each to have the same probability distribution.

In the microconstant case, it was enough to show that the restricted IC-evolution functions maintain a macroperiodic conditional distribution over the effective IC-variables from trial to trial. When trials are not microconstant, if the probabilities of outcomes are to be the same from trial to trial, it seems that the restricted IC-evolution functions will have to maintain exactly the same distribution over the effective IC-values from trial to trial.[2]

Such an effect looks hard to achieve unless the restricted IC-evolution functions are all the same, that is, unless IC-evolution is outcome independent.[3] But this turns out to be an overly pessimistic view. In what follows,

1. But not a necessary condition, as the probability distribution could vary from trial to trial while remaining independent of previous outcomes. Of course, these probabilities could not all equal the relevant strike ratio.

2. This is not a necessary condition: even in a non-microconstant experiment, there may be more than one distribution over the effective IC-variables that yields the same distribution over outcomes. But as far as I know, there is no real world system that exploits this possibility.

3. One might think it is hard to achieve even then, since the distribution to be maintained would have to be a fixed point, so to speak, of the transformation effected by the IC-evolution function. But such a fixed point may well exist, and there may be swift convergence on the point, if IC-evolution is sufficiently inflationary. An inflationary evolution function generates a conditional distribution from a fairly small portion of

I present a less restrictive set of conditions sufficient for the distribution to remain the same from trial to trial. These conditions are still rather tight, but as I will show at the end of this section for the case of kinetic theory, they have a very real and important application in understanding the probabilistic dynamics of complex systems.

The Rippled Wheel

Begin with an example. I deliberately choose an experiment in which the relevant probabilities are not microconstant. Consider an autophagous wheel similar to that described in section 3.74 (all references are to *Bigger than Chaos*), where half the disk is red and half is black and initial velocities are limited to those in the range $0 \ldots v$ that cause the wheel to perform one revolution or less. On this new wheel, however, the velocities are marked on the wheel in such a way that the range $0 \ldots v$ is repeated many times in each colored section, rather than occurring just once. As on the other autophagous wheels I consider in section 3.7, the markings are evenly spaced, but I will suppose, so as to create a case where IC-evolution is not outcome independent, that the range of velocities is repeated more often on the red half of the wheel than on the black half. I call this the *rippled wheel*.

Given a macroperiodic distribution over its initial IC-variable, a chain of trials on the rippled wheel will maintain the same distribution over the effective IC-values whatever outcomes occur. This is the distribution in which the effective IC-value that brings about a red outcome, corresponding to the interval $[0, v/2]$, has the same probability as the effective IC-value that brings about a black outcome, corresponding to the interval $[v/2, v]$, namely, 1/2. (Because initial velocities are limited to those between 0 and v, these are the only two possible effective IC-values for the wheel.)

The reason that the distribution is maintained from trial to trial is, roughly, that the repetition of many $0 \ldots v$ markings means that, although the probability of red or black is not microconstant, the probability of obtaining any given speed as the starting value for the next trial is micro-

the initial distribution. If the initial distribution is fairly smooth, these portions may all be roughly uniform. Take the distribution that the restricted IC-evolution function generates from a uniform distribution. This distribution, if it is itself fairly smooth, will be the desired fixed point.

constant. As a result, because the rippled wheel satisfies the appropriate version of the sufficient conditions for the independence of microconstant probabilities in chained trials—the conditions stated in section 3.76—the probability distribution over the initial speeds for each trial, conditional on all previous speeds, is the same, as desired.

More formally, chained trials on the rippled wheel are independent because (a) the wheel's IC-evolution function is such as to impose a single microconstant probability distribution over the effective IC-values for each trial in a chain,[4] and (b) the wheel satisfies the sufficient conditions for the independence of chained microconstant trials. What I will do in the remainder of this section is discuss the reasons that (a) is true, that is, the reasons that the initial speed distribution for a rippled wheel is microconstant. My aim is to state the reasons as generally as possible, so as not only to explain independence on the rippled wheel, but also to give sufficient conditions for chained non-microconstant trials generally to have effective IC-variables with microconstant distributions.

The reason for the microconstancy is that IC-evolution on the rippled wheel has the following properties:

1. It is *folding*, and
2. It is *well-tempered* relative to the effective IC-values for the succeeding trial.

In what follows I state the nature of these properties, describe how they lead to a microconstant effective IC-variable distribution, and thus show how they are conducive to the maintenance of a unique conditional probability distribution over the effective IC-values in a chain from trial to trial.

The Folding IC-Evolution Function

The notion of folding appeared in section 3.73 as an aspect of the stretch-and-fold action of the IC-evolution function for the straight wheel. Folding turned out not to be important for the microconstant case, but it is a key ingredient in the non-microconstant case. The kind of folding required

4. Note that I talk about the effective IC-variable distribution here, whereas in the previous paragraph I talked about the full IC-variable distribution. This is the first step in the weakening, and therefore the generalizing, of the sufficient conditions.

is as follows: the interval corresponding to each effective IC-value for a trial on the experiment must be divisible into many contiguous parts, each of which parts is mapped by the restricted IC-evolution function for that IC-value onto exactly the same range of effective IC-values for the next trial.[5] The effective IC-values in this range, then, will be the only possible effective IC-values for any trial in the chain.[6] In the language of stretch and fold, each restricted IC-evolution function must stretch out the interval corresponding to its effective IC-value considerably, and then fold the stretched interval neatly into a given range of effective IC-values. The folding requirement, then, implies a considerable degree of inflation. When the restricted IC-evolution functions satisfy this requirement, I say that the entire IC-evolution function is *folding* (see figure 3.7A.1).

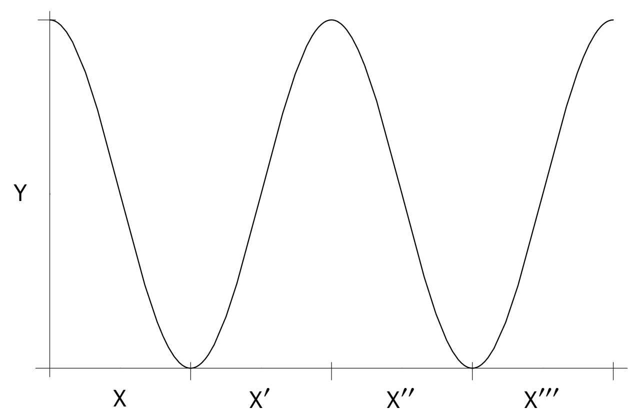

Figure 3.7A.1: A function is folding if its domain can be divided into contiguous sets (X, X', X'', and so on in the figure) each of which maps onto the same range of values (Y in the figure).

The rippled wheel is folding because the *same range* of initial speeds $0 \ldots v$ is embedded *many times* in each colored section of the wheel.[7] An interval corresponding to a single effective IC-value for the wheel—a

5. This is folding in a loose sense, then; see note 24, chapter three of *Bigger than Chaos*.

6. Except the first; see below.

7. Each restricted IC-evolution function of the rippled wheel, then, not only has the same range of effective IC-values but the same range of full IC-values. This strictly stronger property is not necessary for a partition to qualify as a buffer partition. *Buffer partition* is defined later in this section.

maximal contiguous interval of speeds all causing the pointer to indicate the same colored sector—can be divided into as many parts as there are repetitions of $0 \ldots v$ in that section. Each part will then map, by way of the relevant restricted IC-evolution function, to the range of speeds $0 \ldots v$, as mandated by the folding requirement.

Let me continue with the help of an example. Consider an effective IC-value for a spin on the rippled wheel, say, a value giving rise to a red outcome. Call it z. Given my earlier assumptions, that all speeds fall into the range $0 \ldots v$ and that the maximum speed v causes the wheel to complete just one full revolution, z will correspond to fully one half of the range $0 \ldots v$, say, $[0, v/2]$. Suppose that the range $0 \ldots v$ is repeated 20 times in the red half of the wheel. Then the interval $[0, v/2]$ corresponding to z can be partitioned into 20 contiguous sets, each of which is mapped by the restricted IC-evolution function to the full range of spin speeds $0 \ldots v$. For reasons that I about to explain, I call this partition z's *buffer partition*. The folding requirement, then, guarantees the existence of a many-membered buffer partition for each effective IC-value.

Folding plays the following role in the creation of a microconstant probability distribution over an experiment's effective IC-variables: it delineates the optimal constant ratio partition for the microconstant evolution function. That partition, for a given restricted IC-evolution function, will simply be the corresponding effective IC-value's smallest buffer partition.[8] From this point on, when I talk of *the* buffer partition for an effective IC-value, I mean its smallest buffer partition.

If the buffer partition for an effective IC-value is to serve as a constant ratio partition for the corresponding restricted IC-evolution function, it must satisfy the following two conditions:

1. The buffer partition must have many members, each including in its range every possible effective IC-value, and
2. The strike ratio for each possible effective IC-value must be the same within each member of the buffer partition.

Condition (1) is guaranteed by the folding requirement. In order to guarantee condition (2), a second requirement must be introduced; this is the requirement that IC-evolution be well-tempered, to be discussed shortly.

8. If there is more than one equally small buffer partition, there is more than one optimal constant ratio partition.

A buffer partition, as defined above, is a partition of the interval corresponding to a single effective IC-value. But in what follows, I will use the term chiefly to mean the union of the buffer partitions for all of an experiment's possible effective IC-values, which constitutes a partition of the complete range of the experiment's IC-variable.

I will now explain the reason for the term *buffer partition* and, in so doing, help the reader to see more clearly the role of folding in the treatment of non-microconstant experiments. This discussion introduces a way of thinking about the rippled wheel that does not explicitly invoke the microconstancy of the effective IC-value distributions and that follows more closely the formal results relevant to the independence of non-microconstant trials stated in sections 3.B2 and 3.B4; I hope the contrast will be illuminating rather than confusing, but the reader may safely move on to the discussion of well-temperedness.

Consider the way in which the full IC-value for the i^{th} trial on a rippled wheel determines the full IC-value for the $(i+1)^{\text{th}}$ trial. The process can be thought of as having two parts. First, the high level information in the IC-value chooses a member of the experiment's buffer partition. Second, the rest of the information chooses a particular point in that particular member of the buffer partition, and in so doing, fixes the full IC-value for the $(i+1)^{\text{th}}$ trial. Now divide the full IC-value for the i^{th} trial into two parts, the first part corresponding to the information that determines the buffer partition member, the second part corresponding to the rest of the information in the IC-value. Call the higher level information the *buffer information*; the buffer information, then, is just the information that determines into which buffer partition member the IC-value falls.

Why the term *buffer*? The difficulty posed by outcome dependent IC-evolution is that the conditional probability distribution over the effective IC-value for the $(i+1)^{\text{th}}$ trial will depend on the effective IC-value, hence quite possibly on the outcome, of the i^{th} trial, since different restricted IC-evolution functions will transform the distribution for the i^{th} trial in different ways. In the case where trials were microconstant, this was not a problem provided that the different distributions produced by the restricted IC-evolution functions were all macroperiodic. But the difference between different macroperiodic distributions matters in the non-microconstant case.

The role of buffer information is to soak up, in a sense, the difference between different macroperiodic distributions, as follows. The fundamental fact on which, in a way, all of chapters two and three of *Bigger than Chaos* is based, is that what differentiates any two macroperiodic distributions is only the way in which high level information about their variables is correlated (see, in particular, sections 3.43 and 3.5, and example 3.5 of section 3.B4). In an experiment with folding IC-evolution, this high level information is all buffer information. But folding IC-evolution by definition discards buffer information. As a result, the differences between macroperiodic distributions are discarded, and so these differences make no difference to the distribution over the full IC-values for the next trial.

Let me put this more precisely. Provided that the distribution over the full IC-values for the experiment is macroperiodic relative to the buffer partition, the distribution over the low level information obtained by discarding the buffer information will be approximately uniform (theorem 3.10). That is, the probability distribution over the relative position of a full IC-value inside its particular member of the buffer partition will be uniform. Then, provided that a uniform distribution over relative position within any buffer partition member leads to the same distribution over effective IC-values for the subsequent trial, the conditional effective IC-variable distributions for different trials in a chain will be independent. Or, restating the proviso in a more familiar way: provided that the strike ratio for any effective IC-value is the same in each member of the buffer partition, independence holds. The well-temperedness requirement is introduced to guarantee that the proviso holds.

The Well-Tempered IC-Evolution Function

In order to derive the microconstancy of the effective IC-value distributions for the rippled wheel, what remains is to state some condition guaranteeing that the strike ratio for any effective IC-value within each member of the buffer partition is the same. The definition of a well-tempered IC-evolution function makes this guarantee in the most direct way possible: I say that a folding IC-evolution function is *well-tempered* relative to a set of outcomes just in case the strike ratios for those outcomes are the same for every member of the experiment's buffer partition.[9] Then what is required is that the IC-evolution function is well-tempered with respect to the events

9. There is an important parallel between well-tempered IC-evolution and well-tempered information (for well-tempered information, see definition 3.15 in *Bigger than Chaos*). The parallel is made clear in note 11.

of obtaining each of the possible effective IC-values for the next trial.[10]

More formally, for every effective IC-value z there must exist some a_z such that for every member U of the buffer partition, the proportion of full IC-values in U that produce an effective IC-value equal to z for the next trial is a_z.

The rippled wheel's IC-evolution is well-tempered because the speed markings are distributed the same way within each repetition of $0 \ldots v$, within a constant of proportionality. More generally, IC-evolution will be well-tempered on those experiments in which the IC-evolution of an IC-value is entirely determined by the relative position of that value within its part of the buffer partition (this is a sufficient condition only); see the comments on kinetic theory at the end of this section for an example.

Now consider as a designated outcome the event of obtaining some particular effective IC-value z on a folding, well-tempered experiment (z, then, is the effective IC-value for the succeeding trial). Take the evolution function for this outcome. The IC-variable space for the experiment can be partitioned into many contiguous sets (because of folding) in each of which the strike ratio for that value is the same (because IC-evolution is well-tempered). The evolution function, then, is microconstant. Provided that a chain begins with macroperiodically distributed IC-values, and provided that IC-evolution maintains the macroperiodic distribution, conditional on any sequence of outcomes—which will be true if the IC-evolution function is sufficiently inflationary and microlinear—the conditional distribution over the effective IC-values will remain the same, as desired.

More formally and more directly, but to the same effect, the following result can be proved: in an experiment with folding, well-tempered IC-evolution, any distribution over the experiment's full IC-values that is macroperiodic relative to the buffer partition for the effective IC-value for the i^{th} trial will induce the same probability distribution over the effective IC-values for the $(i+1)^{\text{th}}$ trial. This result is proved directly as theorem 3.14.[11]

10. That IC-evolution is well-tempered does not entail that it is folding, insofar as folding requires not just the existence of a buffer partition, but the existence of a buffer partition with many members. The definition of well-temperedness allows that the buffer partition may only have one member, but this case is of no interest for the purposes of treating independence in chains.

11. To interpret theorem 3.14 as establishing this conclusion: (a) the partial low level information should have as its level partition the buffer partition for the effective

One remark: observe that it is the macroperiodicity-preserving property of the IC-evolution between the $(i-1)^{\text{th}}$ and the i^{th} trial that determines the conditional probability distribution over the $(i+1)^{\text{th}}$ trial. The IC-evolution between the i^{th} and the $(i+1)^{\text{th}}$ trial also matters, but it is the well-tempered property of this IC-evolutionary step, not its macroperiodicity preservation, that is important. For the non-microconstant chains that satisfy the conditions stated here, then, the fixed distribution over effective IC-values for the i^{th} trial is due not to the macroperiodicity of the full IC-value distribution for the i^{th} trial, but to the macroperiodicity of the distribution for the $(i-1)^{\text{th}}$ trial.

Why, then, should the effective IC-value distribution for the very first trial in such a chain have the same, fixed form? The answer is that it will almost certainly not have this form. In the case of the rippled wheel, for example, the vast majority of macroperiodic distributions over a chain's initial IC-variable will not impose the even distribution over the two effective IC-values, corresponding to the intervals $[0, v/2]$ and $[v/2, v]$, that is the rule in all later trials. Whatever distribution is imposed, however, will be (provided that it is macroperiodic) independent of all the later distributions, which is all that is needed for independence.

It remains only to say how much inflation and how much microlinearity in the restricted IC-evolution functions is sufficient to maintain a distribution over the IC-values that is macroperiodic with respect to the buffer partition. The answers follow immediately from the comments on sufficiency in the treatment of microconstant experiments in section 3.76:

1. Inflation must be sufficient to transform a distribution over any effective IC-value for the i^{th} trial to a distribution over many members of the buffer partition for the $(i+1)^{\text{th}}$ trial. This is already guaranteed by the folding requirement.

2. The evolution function for the i^{th} trial must be microlinear relative to the inverse image of the buffer partition for the $(i+1)^{\text{th}}$ trial.

IC-value of a given trial, say, the i^{th} trial (b) the members of the index set should correspond to possible effective IC-values for the $(i+1)^{\text{th}}$ experiment in the chain, (c) the full index function should therefore be the IC-evolution function for the trial (compare example 3.5 in section 3.B4). The theorem guarantees a fixed probability distribution over the partial low level information conveyed by the index; this is, then, a fixed probability distribution over the effective IC-values for the $(i+1)^{\text{th}}$ trial.

Sufficient Conditions for Independence

Putting everything together, sufficient conditions for the independence of chained trials on a non-microconstant experiment are that

1. The experiment's IC-evolution function is weakly inflationary to a sufficient degree,
2. The experiment's IC-evolution function is sufficiently microlinear,
3. The experiment's IC-evolution function is
   (a) folding, and
   (b) well-tempered relative to the possible effective IC-values of the succeeding trial,
4. The distribution over the chain's initial IC-variable is macroperiodic.

The meaning of *sufficient* in conditions (1) and (2) is spelled out at the end of the last subsection.

To better understand the significance of these conditions, consider how the rippled wheel might be altered without affecting the independence of chained trials. Suppose that the number of times the range of velocities $0 \ldots v$ is inscribed on each section of the wheel is kept constant, so that the folding requirement is satisfied however the wheel is changed. What I want to consider is the question of how different inscriptions of $0 \ldots v$ on the wheel might be individually altered. Suppose that, within any sector of the wheel inscribed $0 \ldots v$, the relative position of a point in the sector is given by a variable x ranging from 0 to 1. The value 0 corresponds to the point at the beginning of the sector (relative to the direction of rotation, say), the value 1 to the point at the end, the value 0.5 to the point exactly halfway between beginning and end, and so on. Then the inscription of the range of velocities $0 \ldots v$ can be encoded as a function f mapping x onto a velocity in the range $0 \ldots v$. On the original rippled wheel, the even spacing of the markings means that all inscriptions are encoded by the function $f(x) = vx$.

The discussion in this section has shown that $f(x)$ can differ from sector to sector. Independence is maintained provided that $f(x)$ is microlinear, to ensure macroperiodicity of the distribution of full IC-values, and provided that $f(x)$ is such as to satisfy the well-temperedness requirement.

When will $f(x)$ make for well-tempered IC-evolution? Let me give an answer just for the case where the probability of red and the probability of black are both 1/2. In this case, $f(x)$ must be such that a uniform distribution over x is equally likely to yield an effective IC-value that will give rise (in the succeeding trial) to a red outcome as to yield an effective IC-value that will give rise to a black outcome. The red and black effective IC-values on the rippled wheel correspond to the intervals $[0, v/2]$ and $[v/2, v]$; the requirement, then, is that an even distribution over x probabilify both equally. This will be true just in case the inverse images, relative to f, of the intervals $[0, v/2]$ and $[v/2, v]$ each have measure 1/2.

There are a number of different ways to meet this requirement. If $f(x)$ is strictly increasing, for example, then the requirement is satisfied if and only if $f(0.5) = v/2$.[12] I leave it to the reader to satisfy themselves that there are many different ways to mark the wheel in accordance with this and the microlinearity requirement.

The Multi-Mechanism Approach

Non-microconstant chains, like microconstant chains, can be treated by taking the multi-mechanism approach. Because the evolution function for effective IC-values is microconstant and IC-evolution is inflationary,

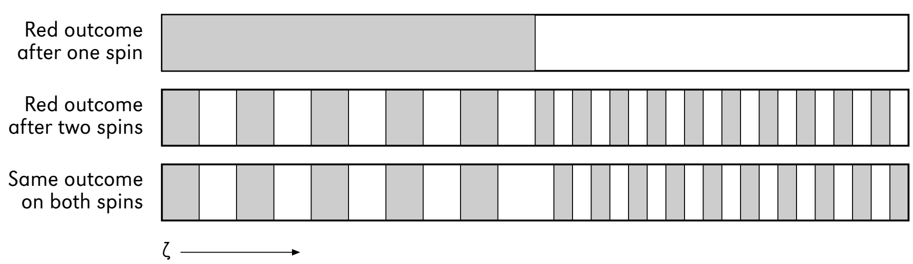

Figure 3.7A.2: Representations of the evolution functions for one- and two-spin experiments on a rippled wheel, together with the evolution function for the composite experiment consisting of trials on one and two spin experiments in the same chain. The designated outcome for the composite experiment is the event of obtaining the same outcome on both trials.

12. I assume, as usual, that f is measure-theoretically well-behaved, to be specific, that it is a Borel function.

the same nesting effect will occur as for microconstant chains, with strike ratios preserved through microlinearity (see sections 3.74 and 3.76). The evolution functions for one- and two-spin experiments on the rippled wheel are shown in figure 3.7A.2; they are clearly *harmonic* in the sense of section 3.74. I assume in the figure that the range $0 \ldots v$ is inscribed five times in the red section of the wheel and ten times in the black section.

An Example from Kinetic Theory

The result derived in this section may be applied to my treatment of kinetic theory in section 4.8 of *Bigger than Chaos*, as follows.

Consider two molecules colliding approximately head-on. Will the first molecule rebound more to the left or more to the right? The answer depends on the molecules' relative position. If the first molecule lies more to the left of the region picked out by its Boltzmann description, a left-leaning rebound is more likely; if to the right, a right-leaning rebound is more likely. Thus the strike ratio for a left rebound increases as the first molecule's position approaches the leftmost point of the region. The corresponding evolution function is therefore not microconstant.

Chained non-microconstant "bounce direction" probabilities may be shown to be independent using this section's results. It will follow that the probability of a left bounce following a molecule's $(i + 1)^{\text{th}}$ collision will be independent of the direction of its i^{th} bounce.

To see this, consider the relative angle of impact for the i^{th} trial (section 4.82) as an effective IC-value for the event of a molecule's bouncing to the left, since it is this angle alone that determines the bounce direction (assuming that left and right are assessed relative to the colliding molecules' relative velocity). As explained in section 4.8, the probability distribution over relative impact angle is microconstant and fulfills the conditions for independence in chained trials. The argument for microconstancy appeals to the same properties that appear in this section under the names *folding* and *well-tempered*. What is left to show is that the IC-evolution of relative impact angle is well-tempered relative in particular to a partition of possible impact angles into two sets, one leading to leftwards and one to rightwards bounces. This follows from the observation that it is an impact angle of zero that separates these two sets, regardless of the shapes of the molecules involved, and that the probability distribution over impact angle is symmetrical around zero.