Evolutionary Theory and the Reality of Macro Probabilities

Elliott Sober
University of Wisconsin-Madison

Evolutionary theory is awash with probabilities. For example, natural selection is said to occur when there is variation in fitness, and fitness is standardly decomposed into two components, viability and fertility, each of which is understood probabilistically. With respect to viability, a fertilized egg is said to have a certain *chance* of surviving to reproductive age; with respect to fertility, an adult is said to have an *expected number* of offspring.[1] There is more to evolutionary theory than the theory of natural selection, and here too one finds probabilistic concepts aplenty. When there is no selection, the theory of neutral evolution says that a gene's chance of eventually reaching fixation is $1/(2N)$, where N is the number of organisms in the generation of the diploid population to which the gene belongs. The evolutionary consequences of mutation are likewise conceptualized in terms of the probability per unit time a gene has of changing from one state to another. The examples just mentioned are all "forward-directed" probabilities; they describe the probability of later events, conditional on earlier events. However, evolutionary theory also uses "backwards probabilities" that describe the probability of a cause conditional on its effects; for example, coalescence theory allows one to calculate the expected number of generations in the past that the genes in the present generation find their most recent common ancestor.

If evolutionary *theory* is inundated with probabilities, is the same true of the *processes* that evolutionary theory seeks to characterize? A straightforward realist interpretation of the theory yields an affirmative answer to this question. Since the theory truly describes what happens in nature, and since the theory describes nature probabilistically, the probabilities it postulates are real. In spite of the simplicity of this interpretation, there have been dissenters. The title of Alexander Rosenberg's (1994) book, *Instrumental Biology or the Disunity of Science*, suggests where he stands on this issue. Rosenberg's thesis is that the probabilities used in evolutionary theory should *not* be interpreted realistically – they are not objective quantities -- because they are mere excuses for our ignorance of detail. For Rosenberg, "evolutionary phenomena are … deterministic, or at least as deterministic as underlying quantum indeterminism will allow (p. 82)."[2] The probabilities of evolutionary *phenomena* are one thing, the probabilities that evolutionary *theory* assigns to those phenomena another.[3]

Although Rosenberg's thesis is about evolutionary theory, his reasons for holding it are general enough to apply to any theory that uses probabilities. And because the motivation for Rosenberg's instrumentalism is so general, it is no surprise that this position was enunciated long before evolutionary theory was mathematized in the 20th century. Rosenberg's thesis traces back to Laplace:

Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, and the past, would be present to its eyes (Laplace 1814, p. 4).

In the *Origin*, Darwin (1859, p. 131) gives voice to the same thought when he explains what he means by saying that variation is "… due to chance. This, of course, is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation."

Laplace was thinking about Newtonian theory when he described his demon, and he took that theory to be deterministic in form.[4] Since that theory makes no use of probabilities, the probabilities we use to describe nature are mere confessions of ignorance. Rosenberg's Laplacean position involves no commitment to determinism. He concedes that if determinism is false, then it is a mistake to claim that the *only* reason we use probabilities to describe nature is that we are ignorant of relevant details. But there is a Laplacean thesis that survives the death of determinism; this is the reductionist idea that *the only objective probability an event has is the one assigned to it by the micro-theory.*[5] If Newtonian theory is the true theory of particles, then no probabilities (other than zero and one) are objective. If Quantum Mechanics is the true theory of particles, then the only objective probabilities are the ones assigned by Quantum Mechanics. Either way, the probabilities assigned by evolutionary theory are not to be interpreted realistically if they differ in value from the ones assigned by whatever the true micro-theory turns out to be.

It is interesting that Laplace says that his demon would find *nothing* uncertain; this goes beyond the more modest claim that the demon would have no uncertainty about the mass, velocity, acceleration, and other properties described by Newtonian theory. Laplace's stronger claim suggests the thought that the properties discussed in Newtonian theory provide a *synchronic supervenience base* for all the other properties that macro-objects might have. And this thought, in turn, can be generalized further, so that there is no reliance on Newtonian theory or on the truth of determinism:

(MS)   A complete specification of the properties that all particles have at a given time uniquely determines *all* the properties that *all* macro-objects have at that same time.

This is the idea of *mereological* (part/whole) supervenience. If two objects and the environments they occupy are particle-for-particle physical copies of each other, they will also be identical in terms of their psychological and biological properties.[6]

The principle of mereological supervenience (MS) says that micro determines macro. It is silent on the converse question of whether macro determines micro. I will assume in what follows that it does not; that is, I'll assume that a given macro-state is *multiply realizable* at the micro-level. The relation of micro to macro is many-to-one. For example, an ideal chamber of gas can have a given temperature (its

---

[4] In fact, there are reasons to think that Newtonian mechanics is not deterministic; see Earman (1986, 2004) and Butterfield (2005) for discussion.
[5] This is the view embraced by the theory of single-case propensities developed by Giere (1973).
[6] (MS) can be modified to accommodate the idea that some macro-properties supervene on historical facts, so that the supervenience base for macro-properties at time t is the state of particles in some temporal interval leading up to t.

mean kinetic energy) by many different assignments of kinetic energy values to its constituent molecules.

The Laplacean picture of a deterministic universe is represented in Figure 1. The MS principle says that a system's micro-state at one time fixes its macro-state at that time; A at t_1 makes it the case that X at t_1, and B at t_2 makes it the case that Y at t_2. This is the meaning of the vertical arrows in Figure 1. The micro-state evolves by deterministic Newtonian rules (represented by the arrow from A to B), so the (complete) micro-state at one time fixes the micro-state at all later times. Since diachronic determination and synchronic supervenience are both necessitation relations, A at time t_1 insures that Y at t_2, by transitivity. If we use probabilities to describe whether macro-state Y will occur at time t_2, given the fact that the system was in macro-state X at t_1, we do so only because of our ignorance. The demon has no need of these probabilities. When the demon predicts the macro-state Y at t_2 from the micro-state A at t_1, it presumably first predicts the micro-state B at t_2 and then figures out that B suffices for the macro-state Y. Laplace's idea thus requires that the demon's knowledge extend beyond Newtonian matters; the demon also needs to be savvy about how Newtonian facts connect to facts described in the vocabularies of other sciences – for example, biology and psychology.

Figure 1: Laplace's Demon in a Deterministic Universe		
	Time t_1	Time t_2
Macro-state	X —[p]→ Y	
Micro-state	A ——→ B	
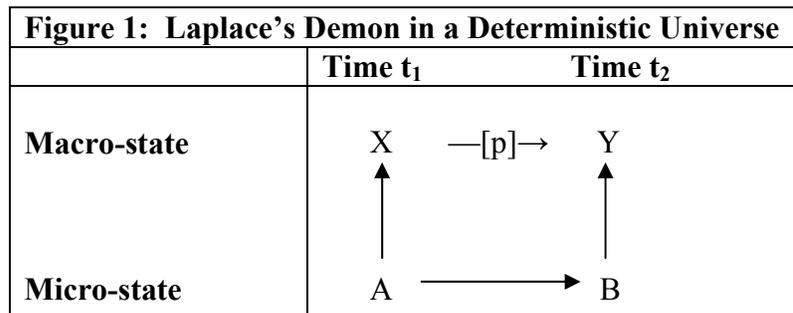
Given this Newtonian description of Laplace's demon, what would the corresponding picture be for a Laplacean who accepts indeterminism as a fact about the physical micro-level?[7] The situation is depicted in Figure 2. If the micro-theory in question is indeterministic, the system's micro-state at t_1 confers probabilities on the different micro-states that might obtain at t_2. Some of these possible micro-states will be supervenience bases for the macro-property Y; others will not. Suppose there are n disjoint micro-states (B_1, B_2, … B_n) that are possible supervenience bases for the macro-state Y and that $Pr(B_i$ at $t_2 \mid A$ at $t_1) = q_i$. We mere mortals, who are aware only of the macro-state X that obtains at time t_1, must predict whether Y will obtain at t_2 by computing the value of $p = Pr(Y$ at $t_2 \mid X$ at $t_1)$. I will call this a *macro-probability* because the conditioning proposition describes the macro-state X that the system occupies at t_1. The demon, who can instantly see which micro-state obtains at t_1, will compute $Pr(Y$ at $t_2 \mid A$ at $t_1)$, which I term a *micro-probability*. Presumably the demon does this first by computing $q = q_1 + q_2 + … + q_n = Pr(B_1$ or B_2 or … or B_n at $t_2 \mid A$ at $t_1)$ and then taking into account the fact that the disjunction (B_1 or B_2 or … or B_n) is equivalent to Y. The point is that the value of $q = Pr(Y$

---

[7] The following remark from Earman (2005) is a useful cautionary reminder to those who think that indeterminism is a settled matter in modern physics: "One might have hoped that this survey [referring to his own paper] would provide an answer to the question: If we believe modern physics, is the world deterministic or not? But there is no simple and clean answer. The theories of modern physics paint many different and seemingly incommensurable pictures of the world; not only is there no unified theory of physics, there is not even agreement on the best route to getting one. And even within a particular theory -- say, Quantum Mechanics or the General Theory of Relativity -- there is no clear verdict. This is a reflection of the fact that determinism is bound up with some of the most important unresolved problems for these theories."

at t_2 | A at t_1) may differ from the value of $p = \Pr(Y$ at t_2 | X at t_1). Although a Laplacean demon in an indeterministic universe will need to use probabilities to predict what will happen,; it will use probabilities that may differ in value from the ones that we less-informed human beings are forced to employ. Were it not for our ignorance, we should do what the demon does, or so the Laplacean claims.[8]

Figure 2: Laplace's Demon in an Indeterministic Universe		
	Time t_1	**Time t_2**
Macro-state	X —[p]→	Y
Micro-state	A —[q]→	(B₁or B₂ or … or Bₙ)
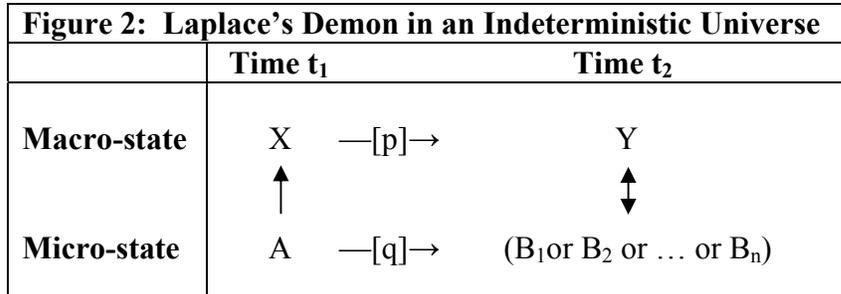
The Laplacean position, then, does not depend on whether determinism is true. The first part of the position is a thesis about prediction:

(L₁)   Suppose you know the system's macro-state (X) at t_1 and also know the system's micro-state (A) at t_1, and you want to predict whether the system will be in state Y at time t_2. If you know the values of both the macro-probability $\Pr(Y$ at t_2 | X at t_1) and the micro-probability $\Pr(Y$ at t_2 | A at t_1), and their values are different, then the micro-probability $\Pr(Y$ at t_2 | A at t_1) is the one you should use to make your prediction.

Read contrapositively, this means that

If you are entitled to use the macro-probability $\Pr(Y$ at t_2 | X at t_1) in predicting whether Y will occur at t_2, this is because $\Pr(Y$ at t_2 | X at t_1) = $\Pr(Y$ at t_2 | A at t_1) or you don't know the value of $\Pr(Y$ at t_2 | A at t_1), or you don't know that A is the micro-state of the system at t_1.

This contrapositive brings out the fact that the Laplacean thinks that there are two possible justifications for using a macro-probability in making a prediction. One involves lack of knowledge; the other is the truth of a probabilistic equality.

The Laplacean principle (L₁) describes the probabilities you should use in making predictions, but does not connect that issue with the question of which probabilities are objective. This further element in the Laplacean position can be formulated as follows:

(L₂)   If the only justification you have (or could have) for using the macro-probability $\Pr(Y$ at t_2 | X at t_1) to predict whether Y will be true at t_2 is that you don't know the value of the micro-probability $\Pr(Y$ at t_2 | A at t_1) or you don't know that A is the micro-state of the system at t_1, then the macro-probability $\Pr(Y$ at t_2 | X at t_1) is not objective.

---

[8] Why does Figure 1 have an arrow going from B at t_2 to Y at t_2, while Figure 2 has a double arrow between (B₁ or B₂ or … or Bₙ) at t_2 and Y at t_2? The reason I drew Figure 2 in this way is that I wanted the demon to be able to derive a point value for $\Pr(Y$ at t_2 | A at t_1) from the fact that $\Pr(B_1$ or B_2 or … or B_n at t_2 | A at t_1) = q. If the disjunction of the B_i at t_2 simply sufficed for Y at t_2, what would follow is just that $\Pr(Y$ at t_2 | A at $t_1) \geq q$.

Whereas (L_1) says that there are two possible reasons for using a macro-probability to make a prediction, (L_2) says that one of those reasons (lack of knowledge) should lead you to attach a subjective interpretation to the macro-probability. Together, these two principles entail the Laplacean thesis that the only way the macro-probability $Pr(Y$ at $t_2 \mid X$ at $t_1)$ can be objective is for it to have the same value as the micro-probability $Pr(Y$ at $t_2 \mid A$ at $t_1)$. If these two probabilities have different values, the macro-probability should not be taken to describe an objective matter of fact.[9]

In what follows, I'll describe what I think is wrong with (L_2); (L_1), as I'll explain, follows from two principles that I'll assume without argument are correct. After criticizing (L_2), I'll try to answer the following two questions: Why should we think that a given macro-probability is objective? And where do objective macro-probabilities come from? My goal is to provide a non-Laplacean account of the epistemology and metaphysics of objective macro-probabilities.

1. The Principle of Total Evidence and Mereological Supervenience

What could motivate the two-part Laplacean position (L_1 and L_2)? It might seem to be an instance of the prejudice that Wilson (2004) calls *smallism* – the idea that it is better to conceptualize the world in terms of parts than in terms of wholes. In fact, the first conjunct in this two-part position (L_1) can be justified in terms of principles that many philosophers find compelling. L_1 follows from the *principle of total evidence* and the *principle of synchronic mereological supervenience*, both of which I'll assume are true for the rest of this paper. The principle of total evidence has nothing explicitly to say about micro and macro; rather, it bids you conditionalize on *all* of the information at your disposal when you try to figure out what the probability is of a future event Y. In particular, if you know both Φ and Ψ, and Φ entails Ψ, then using $Pr(Y \mid \Psi)$ will be a mistake if $Pr(Y \mid \Phi) \neq Pr(Y \mid \Psi)$. Applied to the problem at hand, the principle of total evidence says that if your micro-description A of the state of the system at time t_1 *entails* your macro-description X of the system at that same time, and the two descriptions confer different probabilities on the system's occupying state Y at time t_2, then you should use the former.[10, 11] The (MS) principle completes the argument by affirming the antecedent; it asserts that A at t_1 entails X at t_1, if A is a *complete* description of the system's micro-state at t_1. Notice that this justification of (L_1) does not assert that *every* true micro-description is preferable to *every* true macro-description. If the micro-description is *incomplete*, it may or may not entail the macro-description in question.

---

[9] These two principles, L_1 and L_2, bear slightly different relationships to the question of which interpretation of probability you should use. Whether you are talking about subjective credences or objective chances, you still need advice about which conditional probabilities to use in making predictions. In this sense, L_1 is not wedded to any one interpretation of probability. L_2, however, gives advice about which interpretation of probability you should impose on a given conditional probability.

[10] I am construing the principle of total evidence as saying, not just that you should use *all* of the evidence you have, but that using *more* of the evidence is better than using *less*.

[11] Why accept the principle of total evidence? Good (1967) constructs a decision-theoretic justification. He says that his argument applies to logical and subjective probabilities, but not to frequencies. I take the argument to apply to objective probabilities as well.

2. Must (In)Determinism Percolate Up?

The thesis of synchronic mereological supervenience (MS) places a constraint on how determinism or indeterminism at one level is relevant to the same distinction at another:

If Micro entails Macro and $\Pr(Y$ at $t_2 \mid$ Macro at $t_1) = 1$ (or 0), then $\Pr(Y$ at $t_2 \mid$ Micro at $t_1) = 1$ (or 0).

The reason this principle is true is that the probabilities 1 and 0 are *sticky*. If $\Pr(E \mid \Psi) = 1$ (or 0), then strengthening the conditioning proposition (by substituting Φ for Ψ, where Φ entails Ψ) cannot budge that value.[12]   The principle holds regardless of whether the description of the micro-state is complete. Once again, the contrapositive is interesting:

(P)   If Micro entails Macro and $\Pr(Y$ at $t_2 \mid$ Micro at $t_1) \neq 1$ (or 0), then $\Pr(Y$ at $t_2 \mid$ Macro at $t_1) \neq 1$ (or 0).

Proposition (P) is one way to express the idea that *indeterminism must percolate up*. This percolation principle is a consequence of the axioms of probability; it is not a consequence of those axioms that *determinism* must percolate up.

Notice that the percolation principle (P) has the micro- and the macro-descriptions probabilifying the same proposition, namely Y at t_2. If Quantum Mechanics says that your going to the movies tonight has a probability that is strictly between 0 and 1, then belief/desire psychology cannot assign that event a probability of 0 or 1, if your psychological state supervenes on your quantum mechanical state. Proposition (P) does not assert that if Quantum Mechanics assigns to *micro*-events probabilities that are strictly between 0 and 1, then psychology must assign intermediate probabilities to the *macro*-events it describes. That would be to make the following false claim (where, as before A and B_i are micro-properties and X and Y are macro):

If $\Pr(B_i$ at $t_2 \mid$ A at $t_1) \neq 1$ (or 0), for each $i = 1, 2, \ldots n$, then $\Pr(Y$ at $t_2 \mid$ X at $t_1) \neq 1$ (or 0).

Even if A says that each B_i has an intermediate probability, it may still be true that A says that the disjunction (B_1 or B_2 or … or B_n) has a probability of unity. This is how micro-indeterminism about the relationship of A to each B_i can be compatible with macro-determinism concerning the relation of X and Y. In *this* sense, micro-indeterminism need not percolate up.

If the axioms of probability are *a priori*, as I will assume, then so is (P). The truth of this percolation principle does not depend on anything empirical – for example, on the fact that radiation sometimes causes mutations that change the evolutionary trajectories of populations. Discussion in philosophy of biology of whether micro-indeterminism must percolate up into evolutionary processes

---

[12] The thesis that 1's and 0's are sticky asserts that

If $\Pr(A \mid B) = 1$, then $\Pr(A \mid B\&C) = 1$ (assuming that $\Pr(B)$, $\Pr(B\&C) > 0$).

Proof : If $\Pr(A \mid B) = 1$, then $\Pr(A \mid B\&C)\Pr(C \mid B) + \Pr(A \mid B\&\text{-}C)\Pr(\text{-}C \mid B) = 1$ as well. If $\Pr(C \mid B) > 0$, then $\Pr(A \mid B\&C) = 1$. And $\Pr(C \mid B) > 0$, since $\Pr(B\&C) > 0$, $\Pr(B) > 0$.

has often focused on empirical questions of this sort; see, for example, Brandon and Carson (1996), Glymour (2001), and Stamos (2001). To grasp the difference between these two questions about percolation, consider Figure 3. My question concerns how the relationship of A at t_1 to Y at t_2 constrains the relationship of X at t_1 to Y at t_2, assuming that X supervenes on A. The other question focuses on is whether Q can affect Y by affecting X.

Figure 3: Two Questions about Percolation: (i) If A at t_1 entails X at t_1 and $Pr(Y$ at $t_2 \mid A$ at $t_1)$ is intermediate, must $Pr(Y$ at $t_2 \mid X$ at $t_1)$ also be intermediate? (ii) Can Q at t_0 affect Y at t_2 by affecting X at t_1? Question (i) can be answered *a priori*; (ii) cannot.			
	Time t_0	**Time t_1**	**Time t_2**
Macro-state		X	Y
Micro-state	Q	A	

Although I have interpreted proposition (P) as saying that indeterminism must percolate up, the proposition also asserts that *determinism must filter down.* If we had a true deterministic macro-theory, that would entail that there must be a true deterministic micro-theory. It is interesting that philosophers usually think of micro-theories as constraining macro-theories, but not *vice versa*. Can this asymmetry be justified by pointing to the fact that quantum mechanics provides various no-hidden-variable proofs, whereas there are no deterministic theories in macro-sciences that we are prepared to say are true? This is a question that merits further exploration.

3. Does Macro Screen-off Micro?

If a macro-probability is to pass the Laplacean test for objectivity, the macro-description of the system at t_1 must capture all of the information in the micro-description at t_1 that is relevant to predicting the system's macro-state at t_2. Where X is a macro-property at t_1 and A is a micro-property at that same time, we can express this idea by using Reichenbach's (1956) concept of screening-off:

(Macro SO Micro) $Pr(Y$ at $t_2 \mid X$ at $t_1) = Pr(Y$ at $t_2 \mid X$ at t_1 & A at $t_1)$.[13]

If the micro-state at one time *entails* the macro-state at that time, as the principle of mereological supervenience (MS) asserts, then the micro-state at t_1 screens off the macro-state at t_1 from the macro-state at t_2:

(Micro SO Macro) $Pr(Y$ at $t_2 \mid A$ at $t_1) = Pr(Y$ at $t_2 \mid X$ at t_1 & A at $t_1)$.

These two screening-off principles are compatible.[14] Together they entail that:

---

[13] Strevens (2003) views (Macro SO Micro) as the key to understanding how there can be simple, stable, and objective probabilistic macro-laws; he calls this principle *the probabilistic super-condition*. Strevens is interested in cases in which it is exactly satisfied as well as in cases in which it is satisfied only approximately.

$$\Pr(Y \text{ at } t_2 \mid A \text{ at } t_1) = \Pr(Y \text{ at } t_2 \mid X \text{ at } t_1).$$

This last equality asserts that if you want to predict the system's state at t_2 based on information about its state at t_1, it doesn't matter whether you use the micro- or the macro-description. The two descriptions deliver the same probability.

To see when, if ever, (Macro SO Micro) is true, there are four cases to consider, which are shown in Figure 4. The macro-theory, which relates X at t_1 to Y at t_2, will either be deterministic or indeterministic. I previously have discussed the micro-theory as relating A at t_1 to B at t_2. For purposes of the present discussion, I'm going to understand it as relating A at t_1 to Y at t_2. This theory is "micro" in the sense that it uses the micro-property A to predict whether Y will obtain. The micro-theory, like the macro-theory, will be either deterministic or indeterministic. I assume that A at t_1 suffices for X at t_1; the micro-property is a supervenience base for the macro-property.

Figure 4: When is (Macro SO Micro) true?			
		Macro-theory is	
		Deterministic	Indeterministic
Micro-	Deterministic	Always	Never
Theory is	Indeterministic	Not Defined	Depends on details

Suppose the macro-theory is deterministic (the first column in Figure 4). If so, the micro-theory must be deterministic as well, owing to the stickiness of 1's and 0's; the lower-left hand box in Figure 4 is ruled out. When both theories are deterministic, (Macro SO Micro) is correct. As an example, consider Putnam's (1975) well-known example of the peg and the board. Putnam describes a board that contains two holes; one is round and is 1 inch in diameter; the other is square, and is a little more than 1 inch on each side. The peg he describes is square and is 1 inch on a side. The peg fits through the square hole, but not the round one. Why? Putnam contends that the correct explanation is given by the macro-dimensions just cited. He further claims that a micro-description of the configuration of the molecules in the peg and board is either not an explanation, or is a terrible explanation. Putnam concludes from this that reductionism is false – a macro-story explains something that no micro-story can explain (or explain as well). I prefer a more pluralistic view of explanation, according to which micro- and macro-stories are both explanatory (Sober 1999a); Jackson and Pettit (1992) do too. However, the present point is that, in Putnam's example, the macro-facts about the peg and board screen-off the micro-facts from the fact to be explained. This is because the peg will *necessarily* pass through one hole but not the other, given the macro-dimensions (and the initial condition that the peg is pushed in the right way).

What happens when the macro-theory is probabilistic (the second column in Figure 4)? If the micro-theory is deterministic, the (Macro SO Micro) principle is false. Clearly, if the micro-probability $\Pr(Y \text{ at } t_2 \mid A \text{ at } t_1) = 1$ (or 0), and the macro-probability $\Pr(Y \text{ at } t_2 \mid X \text{ at } t_1)$ has an

---

[14] Note that screening-off, as I define it, is not asymmetric; Micro's screening-off Macro from Y does not preclude Macro's screening-off Micro from Y. Screening-off is often defined as an asymmetric relation. Since my purpose in what follows is to argue that (Macro SO Micro) is rarely, if ever, true, my arguments, which are formulated in terms of the weaker notion, also count against the principle when the stronger interpretation is used.

intermediate value, then $\Pr(Y \text{ at } t_2 \mid X \text{ at } t_1) \neq \Pr(Y \text{ at } t_2 \mid X \text{ at } t_1 \text{ and } A \text{ at } t_1)$.[15] A situation of this type arises when the second law of thermodynamics is formulated probabilistically (saying that the entropy of a closed system at t_2 is, with high probability, no less than the entropy at t_1) and the underlying Newtonian micro-theory is taken to be deterministic. Figure 5 depicts an example described by Albert (2000, pp. 71-73), which he uses to explain ideas due to Zermelo and Loschmidt. A system moves from macro-state A at t_1 to B at t_2 and then to C at t_3, increasing in entropy at every step. A might be an isolated warm room that contains a block of ice, B the same room with a half-melted block of ice and a puddle, and C the room with no ice but a bigger puddle. Suppose that macro-state A happens to be realized by micro-state a, B by b, and C by c, and that the Newtonian laws of motion entail that a necessitates b and b necessitates c. Albert observes that the time-reverse of a, r(a), is a possible realizer of A (he uses the word "compatible"), r(b) a possible realizer of B, and r(c) of C, and that the laws of motion say that a state beginning in r(c) will necessarily move to r(b) and then to r(a). Given this, the macro-state B at t_2 does not screen off its micro-realization b from the macro-state C at t_3. $\Pr(C \text{ at } t_3 \mid B \text{ at } t_2)$ is intermediate, but $\Pr(C \text{ at } t_3 \mid b \text{ at } t_2) = 1$ while $\Pr(C \text{ at } t_3 \mid r(b) \text{ at } t_2) = 0$.

Figure 5: The (probabilistic) second law of thermodynamics says that macro-state A goes, with high probability, to B, and B, with high probability, to C. If A is realized by a, B by b, and C by c, this transition must occur, but if A is realized by r(a), B by r(b), and C by r(c), it cannot.			
	t_1	t_2	t_3
Macro	A —[h]→	B —[h]→	C
Actual Micro-supervenience bases	a —[1]→	b —[1]→	c
Possible Micro-supervenience bases	r(a) —[0]→	r(b) —[0]→	r(c)

The last case to consider is the one in which both the micro- and the macro-theories are indeterministic. In this case, I doubt that there can be a general argument to show that (Macro SO Micro) must *always* be false. However, with one type of exception, I know of no macro-descriptions that screen-off in the way that (Macro SO Micro) requires. The following two examples exemplify very general circumstances in which (Macro SO Micro) is false.

First, consider the macro-statement that $\Pr(\text{lung cancer at } t_2 \mid \text{smoking before } t_1) = x$, where "smoking" means smoking 10,000 cigarettes. Suppose there is a single carcinogenic micro-ingredient A in cigarette smoke and that different cigarettes contain different amounts of that micro-ingredient. This means that smoking 10,000 cigarettes entails inhaling somewhere between g and g+h grams of the carcinogenic ingredient. Given this, the macro-probability $\Pr(\text{lung cancer at } t_2 \mid \text{smoking before } t_1)$ is a

---

[15] Although the equality asserted by (Macro SO Micro) can't be exactly correct when the macro-theory is probabilistic and the micro-theory is deterministic, it can be approximately true. Here's a simple case in which this is so. Suppose that the effect term Y at t_2 involves a statistic concerning outcomes over many trials.   For example, what is the probability that a fair coin will land between 40% and 60% heads when tossed 1000 times? The probability is *nearly* one. The deterministic micro-details (supposing there are such) confer on that outcome a value that is equal to 1. So (Micro SO Macro) is approximately true.

weighted average over the different probabilities of cancer that different levels of exposure to the micro-constituent A induce:

$\Pr(\text{lung cancer at } t_2 \mid \text{smoking before } t_1) =$

$$\int_{n=g}^{n=(g+h)} \Pr(\text{lung cancer at } t_2 \mid \text{inhaling n grams of A before } t_1)\Pr(\text{inhaling n grams of A before } t_1 \mid \text{smoking before } t_1)\,(d\text{n}).$$

If the risk of cancer is an increasing function of the number of grams of A that you inhale, then the macro-probability will have a different value from all, or all but one, of the micro-probabilities:

$\Pr(\text{lung cancer at } t_2 \mid \text{smoking before } t_1) \neq \Pr(\text{lung cancer at } t_2 \mid \text{smoking before } t_1 \,\&\, \text{inhaling n grams of A before } t_1)$ for all, or all but one, value of n.

The point is really very simple: if all the children in a classroom have different heights, at most one of them will have a height that is identical with the average height.[16] [17]

The one kind of case I know of in which (Macro SO Micro) is true where the macro-theory is not deterministic involves a macro-description of the system at t_1 that is *defined* so as to confer a certain probability on a macro-state at t_2. As mentioned earlier, the viability component of an organism's fitness is defined as its probability of surviving from egg (at t_1) to adult (at t_2). This means that

$\Pr(\text{O is alive at } t_2 \mid \text{O is alive at } t_1 \text{ and has a viability} = x) =$
$\Pr(\text{O is alive at } t_2 \mid \text{O is alive at } t_1 \text{ and has a viability} = x \,\&\, \text{O at } t_1 \text{ has genotype G}) = x.$[18, 19]

I take it that the genotypic description is a "micro-description," as compared to the fitness description, which is more macro, since the latter attaches to the whole organism without mentioning its parts. Perhaps a radioactive atom's *half-life* provides a similar example; it is defined so that the screening-off relation holds.

Natural selection might seem to provide the perfect setting for (Macro SO Micro) to be true. It is often claimed that natural selection "cares" only about an organism's phenotype, and not about its genotype; in this vein, Mayr (1963, p. 184), Gould (1980, p. 90), and Brandon (1990) have emphasized the idea that natural selection acts "directly" on phenotypes, and only indirectly on genotypes. Perhaps their point should be formulated in the way that (Macro SO Micro) suggests:

---

[16] The monotonic increase of cancer risk with increase in the number of A particles inhaled is not essential for this argument. For each item in a set to have the same value as the set's average value, the items must all have the same value.

[17] Spirtes and Scheines (2003) discuss a similar example – $\Pr(\text{heart attack} \mid \text{cholesterol level} = \alpha)$ will be an average, since there are different mixtures of high and low density lipids that instantiate the same cholesterol level, and different mixtures confer different probabilities on having a heart attack.

[18] The use of descriptors of the population's state at t_1 that are defined so that they entail a probability distribution for the system's state at t_2 is what leads mathematical models in population biology to have the status of *a priori* mathematical truths (Sober 1984) and to be time-translationally invariant (Sober 1993c).

[19] It might be better to express this screening-off claim by absorbing the facts about viabilities at t_1 into the probability function, thus yielding the equality $\Pr_v(\text{O is alive at } t_2 \mid \text{O is alive at } t_1) = \Pr_v(\text{O is alive at } t_2 \mid \text{O is alive at } t_1 \,\&\, \text{O has genotype G at } t_1)$. That way, the conditional probabilities don't conditionalize on probability statements.

$$\text{Pr(O is alive at } t_2 \mid \text{O has phenotype P at } t_1) = \text{Pr(O is alive at } t_2 \mid \text{O has phenotype P at } t_1 \text{ and genotype G at } t_1).$$

In fact, the temporal gap between t_1 and t_2 provides room for this relationship to be falsified, as Figure 6 illustrates. Consider two individuals (Y and Z) who have different genotypes (G_1 and G_2); they have the same total phenotype (P_1) at t_1 but different chances of surviving until t_2. The reason this is possible is that Y's genotype causes it to develop phenotype P_2 at a time after t_1 but before t_2, while Z's genotype causes it to develop phenotype P_3 ($P_2 \neq P_3$) at that intermediate time. An organism's *entire suite* of phenotypic traits up until t_2 affects its chance of surviving until t_2. Although this entire suite may screen-off genotype from survival (this is a better way to put the Mayr/Gould/Brandon point), the phenotypic traits that an organism has *at a single time* during its development often will not.

	Figure 6: A Failure of Screening-Off. Phenotype at t_1 does not screen off genotype at t_1 from being alive at t_2.					
	Organism Y			**Organism Z**		
	t_0	t_1	t_2	t_0	t_1	t_2
Macro		$P_1 \longrightarrow$ Y is alive P_2			$P_1 \longrightarrow$ Z is alive P_3	
Micro	$G_1 \longrightarrow G_1$			$G_2 \longrightarrow G_2$		
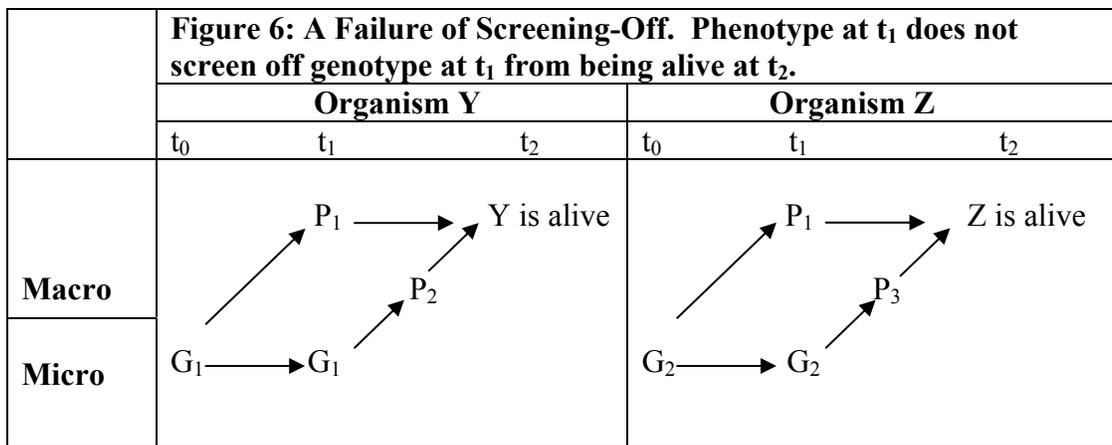
The arrows in Figure 6 all represent causal relations. I assume that an organism's genome is completely stable throughout its lifetime, though this isn't essential for my point; allowing for mutation would mean that an organism's genome at one time exerts a probabilistic (not a deterministic) influence on its genome later. I also assume that cause must precede effect; this is why no arrow connects the genotype at t_1 to the phenotype at t_1. However, the argument against (Macro SO Micro) does not depend on ruling out simultaneous causation. It also does not matter whether genotype at t_0 causally determines phenotype at t_1; this means that even if genotype at t_1 provided a supervenience base for phenotype at t_1, there still would be a counterexample to (Macro SO Micro).

Figure 6 depicts a general circumstance in which (Macro SO Micro) fails. The reason the macro-state at t_1 does not screen-off the micro-state at t_1 from the macro-state at t_2 is that there are *two causal pathways* from the micro-state at t_0 to the macro-state at t_2. The macro-state at t_1 occurs on just one of them.[20] A predictor, demonic or human, who wants to say whether an organism will be alive at

---

[20] The (Macro SO Micro) principle would not be rescued by allowing the times t_0, t_1, t_2 to be intervals rather than instants. As long as there is a temporal gap between t_1 and t_2, the kind of counterexample contemplated here can arise. Thus, even if the macro-state at t_1 supervenes on "historical" facts at the micro-level that extend back in time from t_1, the counterexample stands. And allowing the phenotype at t_1 to influence the phenotype that arises later won't save (Macro SO Micro), either.

t_2 by using information about the organism's state at t_1 may do better by using the organism's genotype at t_1, rather than the organism's phenotype at t_1, as the basis for the prediction.[21]

A similar argument undermines the claim that phenotype screens-off genotype from reproductive success when an organism's reproductive success is defined as its expected number of *viable* offspring:

E(number of viable offspring that O has │ O has phenotype P) =
E(number of viable offspring that O has │ O has phenotype P & O has genotype G).

For a counterexample, consider a dominant gene A; AA and Aa individuals both have phenotype P, which differs from the phenotype Q that aa individuals possess. Suppose individuals with P have higher viability than individuals with Q. If so, AA individuals will have a greater expectation of viable offspring than do Aa individuals, even though they are phenotypically identical. This is because Aa individuals sometimes have aa offspring, but AA individuals never do (Sober 1992). The two-pathway pattern is present here as well. An individual's genotype influences its own phenotype as well as the genotype its offspring have. This is why screening-off fails.[22]

It is important to bear in mind that my pessimistic evaluation of (Macro SO Micro) is predicated on the assumption that the principle of mereological supervenience (MS) is true. If this assumption is dropped, all four cells in Figure 4 must be reconsidered; for example, indeterministic micro-theories can be consistent with deterministic macro-theories, since the stickiness argument no longer applies. Forster and Kryukov (2003) point out that investigations of the relation of micro- to macro-theories in physics often conceptualize macro-states as probabilistic expectations over possible micro-states; this means that the system's (actual) macro-state at a time is not determined by its (actual) micro-state at that time in any obvious way. Quantum mechanics has forced philosophers to take seriously the possibility that the diachronic thesis of determinism may be false. Perhaps the synchronic determination thesis (MS) should be re-evaluated as well (Crane and Mellor 1990, Sober 1999b).

4. Micro- and Macro-Causation

In the previous section I argued that the micro-state of a system at t_1 is often correlated with its macro-state at t_2, even after you control for the system's macro-state at t_1. This is what it means for the (Macro SO Micro) principle to fail. I now want to argue that this relationship between the micro-state at t_1 and the macro-state at t_2 often isn't a *mere* correlation; the micro-state at t_1 often is a *cause* of the macro-state at t_2.

---

[21] I say "may" because the question of whether G_1 at t_1 or P_1 at t_1 is a better predictor of whether Y is alive at t_2 depends on the quantitative relationships that obtain among the path coefficients we might associate with the arrows in the left-hand diagram in Figure 6.

[22] Although I have argued that (Macro SO Micro) is rarely if ever true when both the macro- and the micro-theories are probabilistic, there may be circumstances in which it is approximately correct. One of the main theses of Strevens (2003) is that (Macro SO Micro) is approximately true when the system is "macro-periodic" and approximately "micro-constant." The former says, roughly, that the probability distribution over initial conditions is smooth; the latter concerns how a given macro-state is realized by various micro-states. It suffices for micro-constancy if all the micro-state realizers of a given macro-state have the same probability of occurring, conditional on the occurrence of that macro-state.

To defend this claim, I want to exploit the suggestive ideas about causation presented by Woodward (2003), who draws on the frameworks developed by Spirtes, Glymour, and Scheines (2000) and Pearl (2000). When two events X and Y are correlated, how are we to discriminate among the following three possibilities: (i) X causes Y; (ii) Y causes X; (iii) X and Y are joint effects of a common cause C?[23] The intuitive idea is that if X causes Y, then intervening on X will be associated with a change in Y; this won't be true if Y causes X, or if X and Y are effects of a common cause C. To make this suggestion precise, one has to define the concept of an intervention very carefully, which Woodward does. An intervention on X with respect to Y causes the state of X to take on a particular value; it therefore cancels the other causal influences that would otherwise impinge on X. In addition, an intervention must be delicate, not ham-fisted; if X and Y are joint effects of a common cause C, then an intervention on X with respect to Y must fix the state of X without simultaneously modifying the state of C. Woodward points out that "intervention" is a causal concept, so the account of causation he gives is not reductive. However, the account is not circular, in that the intervention criterion for when X causes Y does not require that you already know whether X causes Y. Even so, it is a consequence of Woodward's theory that you must have lots of *other* causal knowledge if you are to figure out whether X causes Y.

In the first example discussed in Section 3, smoking and lung cancer are the macro-variables, and the number of grams inhaled of the carcinogenic micro-ingredient A is the micro-variable. It seems clear here that intervening on the micro-variable while holding fixed the fact that someone smokes will change the person's probability of getting lung cancer. This tells you that the micro-state at t_1 causally contributes to the macro-state at t_2. The same pattern obtains in the example depicted in Figure 5 in which phenotype at t_1 and being alive at t_2 are the two macro-variables, and genotype at t_1 is the micro-variable. If we change someone's genotype (from G_1 to G_2, or vice versa), while leaving the phenotype unchanged, the chance of surviving until t_2 will change.

These remarks in favor of the micro-state at t_1's being causally efficacious do not rule out the possibility that the macro-state at t_1 is also causally efficacious (Sober 1999b). An intervention that shifted someone from 10,000 cigarettes smoked to, say, 25, would be associated with a reduction in that person's chance of getting lung cancer. Of course, this drastic reduction in number of cigarettes smoked will entail a drastic reduction in how many A particles are inhaled. And it may also be true that smoking causes cancer only because cigarette smoke contains A particles. But none of this should be taken to refute the claim that smoking causes cancer. There is no conflict between the claim that smoking causes cancer *and* the claim that inhaling A particles causes cancer.[24][25]

---

[23] Woodward accepts the so-called causal Markov principle, which asserts that one of these three possibilities must be true; I do not (Sober 1988, 2001b). Woodward's ideas about intervention are interesting independent of this question.

[24] Kim (1989) argues that behavior cannot have both psychological and neurophysiological causes, since this would imply that behavior is over-determined, a consequence that Kim finds objectionable. However, this is not a case of overdetermination if overdetermination requires only that the two causes each be sufficient for the effect and *independent* (Kim 2000). When Holmes and Watson each shoot Moriarty through the heart at the same time, each cause could occur without the other. But if neurophysiological states (or those states plus relevant facts about the physical environment) provide supervenience bases for psychological states, the two will not be independent. A further objection to Kim's argument is relevant when the causal relation is probabilistic – there is no *over*determination if there is no determination at all.

[25] These brief remarks do not address the question of whether the micro supervenience base should be held fixed when the causal role of a macro-property is tested, or whether the macro-property should be held fixed when the causal role of its supervenience base is assessed; see Sober (1999b) for discussion.

5. The Principle of Total Evidence and Explanation

Suppose we use the macro-probability $Pr(Y$ at $t_2 \mid X$ at $t_1)$ to predict whether Y at t_2 because we know that X is the system's macro-state at t_1 and we don't know the value of the micro-probability $Pr(Y$ at $t_2 \mid A$ at $t_1)$ and also don't know the system's micro-state at t_1. The Laplacean concludes from this that we should give a subjective interpretation to the macro-probability $Pr(Y$ at $t_2 \mid X$ at $t_1)$. We now need to evaluate this piece of advice, embodied in the principle L_2. After all, the Laplacean principle L_1 has to do with which probabilities we should use *to make predictions*. If we use probabilities for other purposes – for example, to construct explanations – perhaps this different venue can provide a reason for thinking that macro-probabilities are objective (Sober 1984).

Does our interest in constructing *good explanations* justify our citing X at t_1 as an explanation of Y at t_2, rather than citing A at t_1 instead, when $Pr(Y$ at $t_2 \mid X$ at $t_1) \neq Pr(Y$ at $t_2 \mid A$ at $t_1)$? To endorse this suggestion, we need not agree with Putnam's claim that the micro-details are explanatorily irrelevant. As argued in the previous section, it is often true that both the micro- and the macro-properties of a system at a given time are causally efficacious. If explanation means causal explanation, then causal explanations can be constructed in terms of micro-properties and also in terms of macro-properties. Is there a sense in which the macro-explanations are objectively better?

Putnam, following Garfinkel (1981), holds that a good explanation will be general, and that more general explanations are objectively better than less general ones. If the micro-property A entails the macro-property X, but not conversely, then X will be true of at least as many systems as A is. This is what Garfinkel and Putnam mean by generality. With respect to Putnam's peg and board, there will be more systems that have the macro-dimensions he describes than have the exact molecular configuration that the micro-description specifies.

My reply is that generality is just one virtue an explanation might have, and that we want different explanations to exhibit different virtues, or different mixes of them. Sometimes we want greater generality; at others we want more causal detail. For example, suppose we are evaluating two explanations of an event; one candidate cites just one of its causes while the other cites two; the second of these will be less general than the first, but will be more detailed. The *desiderata* of breadth and depth of explanation conflict (Jackson and Pettit 1992; Sober 1999) and there is no objective criterion concerning which matters more or that determines what the optimal trade-off between them is.

Even when we focus just on generality, it isn't automatic that macro-explanations trump micro-explanations. First, the Garfinkel-Putnam definition of generality entails that the macro-description of a system at a given will be more general than a micro-description of the system at that time, *when the latter entails the former*. The principle of mereological supervenience (MS) assures us that the two descriptions will be related in this way when the micro-description is *complete*. However, this leaves unsettled which description is more general when the micro-description of the system's state is *not* complete. Garfinkel and Putnam have not shown that a macro-explanation of E is more general than *any* micro-explanation of E, but only that the macro-explanation is more general than a *complete* micro-

explanation. Their argument against reductionism therefore fails, even if we accept their definition of generality. The second problem has to do with Garfinkel and Putnam's decision about how generality should be defined. Consider the causal relationships depicted in Figure 7. The distal cause D causes each of P_1, P_2, …, P_n, and each of these proximate causes has its own effect E_1, E_2, …, E_n. Let's suppose that D suffices, but is not necessary, for P_2. The Garfinkel-Putnam definition of generality will then say that P_2 is a more general explanation of E_2 than D is. Their definition of generality focuses on the *explanandum* E_2 and asks how many systems that have this target property have D and how many have P_2. However, there is another definition of generality, one which focuses on how many phenomena a given *explanans* explains. D explains all of the P_i, and all of the E_i as well, whereas P_2 explains only E_2. In this sense, D is more general than P_2 (Tsai 2004). There is no conflict here, of course; P_2 is a more general explanation than D is *of a single explanandum* (namely E_2), but D applies to more *explananda* than P_2 does. The problem is that the Garfinkel-Putnam argument depends on using one definition of generality and ignoring the other.[26]

Figure 7: P_2 may be a more general explanation of E_2 than D is, but D explains more phenomena than P_2 does.
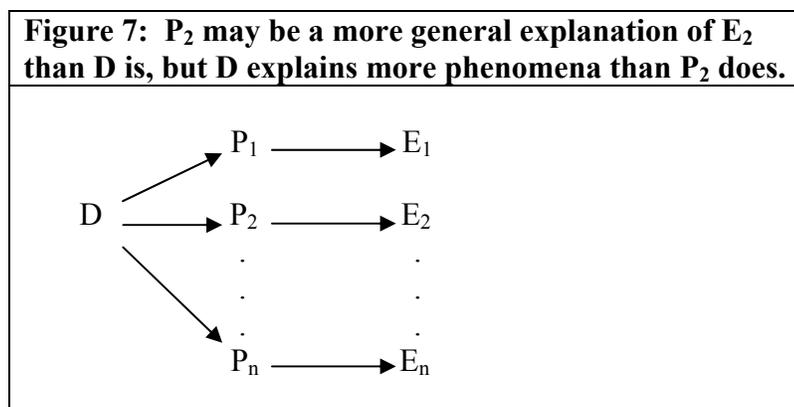
The Laplacean principle L_1 is correct as a claim about prediction, assuming as I do that the principle of mereological supervenience (MS) and the principle of total evidence are both correct. However, the objectivity of macro-probabilities is not to be secured by pointing out that one of our goals in constructing explanations is that they be general. Sometimes we want our explanations to be more general; at other times we want them to be more detailed. But in both cases, we want the facts we cite to be facts – we want them to be objectively correct. To echo the *Euthyphro*, our citing a macro-probability when we give an explanation is not what makes that probability objective; rather, its objectivity is a requirement we impose on the items we choose to include in our explanations. So even if generality of explanation were a categorical imperative, that would not show that macro-probabilities are objective. And, in any event, generality is just one explanatory virtue among several.

6. The Principle of Total Evidence, Objectivity, and the Smart Martian Problem

---

[26] It may be objected that D and P_2 do not compete with each other as explanations of E_2 and that the *desideratum* of generality is relevant only when the task is to sort out competing explanations. If competing explanations must be incompatible, then I agree that "D occurs" and "P_2 occurs" are not competitors (and the same holds for "D causes E_2" and "P_2 causes E_2"); but then it follows that the micro- and macro-stories that Putnam describes in the peg-and-board example are compatible, so they are not competitors, either. On the other hand, if competition does not require incompatibility, I am unsure how this objection should be interpreted.

The Laplacean argument against the objectivity of macro-probabilities resembles an argument that Nozick constructed (described in Dennett 1987) concerning the reality of beliefs and desires. Nozick pointed out that smart Martians will be able to predict our behavior without needing to attribute beliefs and desires to us. They can grasp at a glance the properties of the elementary particles that make up our bodies and use that description as an input to a dynamical physical model to predict how our bodies will comport themselves in the future. Nozick took this to raise the question of why we should think that people really do have beliefs and desires, a question that Dennett (1987) answered, following the Garfkinkel-Putnam line, by citing our penchant for constructing general explanations. The assumption behind Nozick's puzzle is that the only reason there could be for thinking that individuals have beliefs and desires is that we need to postulate those states to predict behavior.[27] We mere mortals need to attribute beliefs and desires if we wish to predict a person's behavior, but smart Martians do not. This seems to rob the existence of beliefs and desires of their objectivity.

The main thing wrong with this argument is its *ad hominem* quality. If you want to know whether something exists, you should ask to see the relevant evidence. Whether you or anyone else *needs* to believe that the thing exists for purposes of prediction (or explanation) is not separately relevant. Nozick's demon won't need to say that bowling balls exist, but that hardly shows that there are none (Sober 1999b). The same holds for the existence of beliefs and desires, and also for objective macro-probabilities.[28]

The comparison that Nozick invites us to make is between two hypotheses, one ascribing a set of beliefs and desires to someone, the other ascribing some complex micro-physical state to that person. According to Nozick's story, the Martian knows that the second of these hypotheses is true, and so he has no need to consider the first. This fact is supposed to raise the question of why we should think that people really have beliefs and desires. Why shouldn't we treat this posit instrumentally, as a fiction that we find useful? However, if the question is whether a person really has some set of beliefs and desires, the appropriate alternative hypothesis to consider is not that they occupy some complex micro-physical state. These are not competing hypotheses – both could well be true. To treat them as competitors would be like wondering whether someone has smoked cigarettes, and then taking the alternative to be that he has inhaled carcinogenic A particles.

If the Laplacean argument against the reality of macro-probabilities were sound, it would be possible to strengthen it. Consider a hypothetical being that has perfect precognition. Unlike Laplace's demon, it doesn't need to observe the present state of the universe and then compute its future; this being knows the whole history -- past present, and future – directly. This super-demon would not need to use any dynamical law -- micro or macro, deterministic or indeterministic -- to predict the future. But surely this does not show that dynamical laws (e.g., those of Quantum Mechanics) are never objectively true – that they are just useful fictions. Here again, what a hypothetical demon needs does not settle what is objectively true.

---

[27] This part of Nozick's argument derives from the indispensability arguments used by Quine (1953) and Putnam (1971) concerning the reality of mathematical entities. For a critique of these indispensability arguments, see Sober (1993a, 2000).

[28] I am not arguing that the existence of beliefs and desires (and of bowling balls, for that matter) is beyond dispute. Perhaps Patricia and Paul Churchland are right that beliefs and desires don't belong in a scientific psychology. My point is that this eliminativist conclusion is not supported by the fact that a hypothetical demon wouldn't need to postulate such things in order to make predictions.

The mistaken idea that the Laplacean criterion (L_2) is a good test for whether a probability is objective may seem to receive support from the equally mistaken assumption that a proposition has just one true probability. This second mistake leads to the pseudo-problem of trying to figure out what that one true probability is. The grip of this misconception can be broken by taking conditional probability, rather than unconditional probability, as one's fundamental concept (Hajek 2003).[29] The proposition that Y occurs at t_2 has a different probability, depending on which conditioning proposition is chosen. To ask whether $Pr(Y$ at $t_2 \mid X$ at $t_1)$ or $Pr(Y$ at $t_2 \mid A$ at $t_1)$ is the true probability of Y at t_2 is like asking what the true distance is to Madison, the distance from Baltimore to Madison or the distance from Boston to Madison.[30]

7. The Epistemology and Metaphysics of Objective Macro-Probabilities

If the needs of demons are not relevant to finding out whether a probability is objective, where should we look? We may begin by asking what it means to say that the macro-probability $Pr(Y$ at $t_2 \mid X$ at $t_1)$ or the micro-probability $Pr(Y$ at $t_2 \mid A$ at $t_1)$ describes objective features of the world. Under the heading of "objective interpretations of probability," there is a now-familiar list – actual relative frequency, hypothetical relative frequency, and propensity, each with variations on those broad themes (Eells 1981). If *any* of these interpretations is correct for *any* statement that assigns a value to a macro-probability, the objectivity of that statement is vouchsafed. For example, if Pr(this coin lands heads $\mid$ this coin is tossed) is taken to describe the actual frequency of heads in a run of tosses, there is nothing subjective about a statement that assigns this probability a value. It is perfectly possible for the statement to be objectively true.[31]

Unfortunately, this does not suffice to show that macro-probabilities are objective, since none of these interpretations is adequate as a general account of objective probability. The objections I have in mind are familiar. With respect to the actual frequency interpretation, the fact is that we often conceptualize probabilities in such a way that they can have values that differ from actual frequencies; for example, a fair coin can be tossed an odd number of times and then destroyed. The other objective interpretations fare no better. If propensities are causal tendencies – that is, if $Pr(Y$ at $t_2 \mid X$ at $t_1)$ represents the causal tendency of X at t_1 to produce Y at t_2 – then the propensity interpretation cannot make sense of the "backwards probabilities" mentioned at the start of this paper that have the form $Pr(X$ at $t_1 \mid Y$ at $t_2)$, at least not if cause must precede effect (this objection is due to Paul Humphreys; see

---

[29] Here are two reasons for thinking that conditional probability should not be defined in terms of unconditional probabilities. First, there are cases in which it is perfectly clear that $Pr(Y \mid X)$ has a value even though $Pr(X)=0$. Second, one sometimes knows perfectly well the value of $Pr(O \mid H)$ and that this is an objective quantity, even though it seems perfectly clear that $Pr(H)$ has no objective value at all; this situation often arises when H is a large-scale scientific theory (e.g., relativity theory), and O is some observational claim upon which the theory confers a probability. These points do not undercut the fact that $Pr(Y \mid X) = Pr(Y\&X)/Pr(X)$ when the two unconditional probabilities are well-defined.

[30] As further evidence against the claim that a proposition has one true probability, it is worth noting that, in an indeterministic process, the probability of Y at t will *evolve* (Sober 1984). For example, it is perfectly possible that $Pr(Y$ at $t_2 \mid A$ at $t_1) \neq Pr(Y$ at $t_2 \mid B$ at $t_0)$.

[31] This point about the actual relative frequency interpretation reveals a confusion in the Laplacean position. The fact that we shouldn't use $Pr(Y$ at $t_2 \mid X$ at $t_1)$ to predict whether Y at t_2 if we can use $Pr(Y$ at $t_2 \mid A$ at $t_1)$ instead doesn't show that the former has no objective interpretation. That conflates a question about pragmatics (which probabilities we should *use* to perform which tasks) with a question about semantics (Sober 2003a).

Salmon 1984, p. 205).[32] On the other hand, if the concept of propensity is stripped of this causal meaning, it isn't clear how the propensity interpretation helps clarify the concept of objective probability. As for the hypothetical relative frequency interpretation, it overstates the relation of probability to what would happen in the infinite long run. It is possible for a fair coin to land heads each time in an infinite run of tosses (though this, like all the other exact sequences that can occur, has a probability of 0). The coin's probability of landing heads is *probabilistically* (not *deductively*) linked to what would happen in the long run, finite or infinite (Skyrms 1980); the hypothetical relative frequency interpretation therefore does not provide a reductive definition of objective probability.

In view of the failures of these interpretations, my preference is to adopt a *no-theory theory of probability*, which asserts that objective probability is not reducible to anything else. Frequencies provide evidence about the values of probabilities, and probabilities make (probabilistic) predictions about frequencies, but probabilities don't reduce to frequencies (Levi and Morgenbesser 1964; Levi 1967; Sober 1993b, 2003b). Instead, we should view objective probabilities as theoretical quantities. With the demise of logical positivism, philosophers abandoned the idea that theoretical magnitudes such as mass and charge can be reduced to observational concepts that are theory-neutral. We should take the same view of objective probabilities.

If we reject the need for a reductive interpretation of objective probability, what does it mean to say that a probability is objective? Taking our lead from other theoretical concepts, we can ask what it means to say that mass is an objective property. The idea here is that mass is a mind-independent property; what mass an object has does not depend on anyone's beliefs or state of mind.[33] The type of independence involved here is conceptual, not causal – it is not ruled out that an object have the mass it does because of someone's beliefs and desires. The next question we need to ask is epistemological -- what justifies us in thinking that mass is an objective property? If different measurement procedures, independently put to work by different individuals, all lead to the same estimate of an object's mass, that is evidence that mass is an objective property. The matching of the estimates is evidence that they trace back to a common cause that is "in" the object; since the estimates do not vary, differences among the procedures used, and among the psychological states of the investigators, evidently made no difference in the results obtained. Of course, this sort of convergence does not *prove* that there is an objective quantity that the observers are separately measuring.[34] It is possible that the observers have some psychological property in common, and it is this subjective commonality that causes their estimates to agree, there being no mind-independent reality that they are bumping up against at all.[35]

---

[32] In addition, $\Pr(Y$ at $t_2 \mid X$ at $t_1)$ will fail to represent the propensity of X at t_1 to cause Y at t_2 when Y at t_2 and X at t_1 are effects of a common cause at t_0. And even when X at t_1 has a causal impact on Y at t_2, the measure of this impact is not to be found in the value of $\Pr(Y$ at $t_2 \mid X$ at $t_1)$; the effect of X at t_1 on Y at t_2 is to be found by making some sort of comparison – e.g., between $\Pr(Y$ at $t_2 \mid X$ at $t_1)$ and $\Pr(Y$ at $t_2 \mid notX$ at $t_1)$ when one controls for the other causally relevant properties of the situation at t_1. Causes are difference makers, so a single conditional probability does not represent a causal propensity.

[33] This definition of objectivity has the consequence that statements about someone's psychological state cannot be objectively true. To handle this type of case, it would be better to define objectivity by saying that believing the statement does not *make* it true (though puzzle cases would still remain, such as the proposition "I believe something"). These niceties won't matter for what follows.

[34] There is a difference between showing that our beliefs about mass are caused by something in the object and showing that the property of the object that does the causing is the object's *mass*. Because of space limitations, I'll glide over this distinction.

[35] Here I am adapting the common cause argument for realism that Salmon (1984) and Hacking (////) defend. I do not claim that the anti-realist has no reply; rather, my point is that macro-probabilities are in the same boat as other theoretical

A perfect matching of the estimates that different investigators obtain is not necessary for the common-cause argument I am describing to go forward. Suppose the different estimates differ, but only a little. That too would support an argument for there being an objective quantity that the different investigators are measuring. Put formally, this amounts to endorsing a model that says that there exists a true value of the mass of the object in question, and that the observers obtain their estimates by processes that are characterized by a set of error probabilities. Endorsement of this model should be understood as a comparative claim, not an absolute claim – we are judging that this model is *better* than one or more alternatives. The alternative of interest here is a model that says that each investigator is measuring a separate property of his or her state of mind. This more complex model can be made to fit the data, but its greater complexity counts against it. In this way, the question of whether mass is objective can be turned into a problem of model selection.[36]

How can we use the example of mass to guide our thinking about the objectivity of macro-probabilities? To begin with, we must take account of a difference between mass and probability. Mass is a non-relational, intrinsic property of an object, but, as mentioned earlier, I want to regard conditional probability, not unconditional probability, as the fundamental notion, and conditional probability is a relation between pairs of propositions. The case for the objectivity of macro-probabilities is not defeated by the fact that you and I will give different probabilities for a coin's landing heads if we conditionalize on different information about the system's initial conditions. It is obvious that $\Pr(Y \mid X)$ and $\Pr(Y \mid A)$ can have different values. Our question is whether $\Pr(Y \mid X)$ and $\Pr(Y \mid A)$ can both be objective quantities.

It may seem that focusing on conditional probabilities does not change matters. If you have background knowledge U and I have background knowledge M, then you will evaluate $\Pr(Y \mid X)$ by computing $\Pr(Y \mid X\&U)$, while I'll evaluate $\Pr(Y \mid X)$ by computing $\Pr(Y \mid X\&M)$. If $\Pr(Y \mid X\&U) \neq \Pr(Y \mid X\&M)$, you and I will assign different values to $\Pr(Y \mid X)$. This suggests that $\Pr(Y \mid X)$ does not represent an objective relation that connects the propositions Y and X; rather, there is a third *relatum* that is not explicitly mentioned, and this is the set of beliefs that some agent or other possesses. However, if this is the right take on $\Pr(Y \mid X)$, what should we say about $\Pr(Y \mid X\&U)$ and $\Pr(Y \mid X\&M)$? Do *they* have unique values, or are their values again relative to the background beliefs of some agent? If the values aren't unique, how did you and I manage to assign values to $\Pr(Y \mid X)$ by consulting our background knowledge U and M and then deciding what the values are for $\Pr(Y \mid X\&U)$ and $\Pr(Y \mid X\&M)$? In fact, what gets added to X and Y when U and M are taken into account is not anyone's psychological state, but certain further propositions. It isn't *your believing* U that matters, but simply the proposition U that you believe; this is what gets included as one of the conditioning propositions in $\Pr(Y \mid X\&U)$. These considerations suggest the following thesis: *some* conditional probabilities have unique values, which are not relative to anyone's background beliefs. This claim is of

---

quantities, like mass. This should be enough to show that there is no *special* objection to thinking that macro-probabilities are objective.

[36] AIC and other criteria of model selection impose penalties on a model for its complexity; see Burnham and Anderson (1998) for discussion. Model selection criteria permit one to answer the question of how much variation in the estimates made by different observers is consistent with regarding the common-cause model as better than the alternative model that postulates separate causes, one for each observer. The common-cause argument given here does not require a commitment concerning which model-selection criterion is best.

course compatible with the concession that *many* conditional probabilities do not have this status.[37] Statisticians have a term for probabilities of the first sort – these are the probabilities that figure in "simple" (as opposed to "composite") statistical hypotheses.

To develop my argument for the reality of (some) macro-probabilities, I want to consider a Newtonian model of coin tossing due to Keller (1986) and Diaconis (1998). The initial conditions for a toss determine whether the coin will land heads or tails. The reason a coin exhibits some mixture of heads and tails in a series of tosses is that the initial conditions vary from toss to toss. To simplify matters, we assume that there is no air resistance, that the coin spins around a line through its plane, and that the coin lands without bouncing (perhaps in sand or in your hand). The relevant initial conditions are then fixed by specifying the values of V (the upward velocity of the toss) and ω (the angular velocity, given in revolutions per second). If V is very low, the coin doesn't go up much; if ω is very low, the coin, as Diaconis puts it, "rises like a pizza without turning over." Depending on the values of V and ω, the coin will turn 0, 1, 2, 3, … times before it lands. Suppose the coin we are considering starts each tossing session by being heads up in the tosser's hand. Then, if the coin turns over 0 or an even number of times, it lands heads, and if it turns over an odd number of times, it lands tails. These different possibilities correspond to the regions of parameter space depicted in Figure 8. Starting at the origin and moving Northeast, the different stripes correspond to 0 turns, 1 turn, 2 turns, etc. For the purposes of investigating how macro-probabilities can be objective in a deterministic universe, I'm going to assume that the Keller/Diaconis model is true.[38] Laplace's demon, since it knows the exact values of V and ω that characterize a given toss, will be able to predict whether the coin will land heads or tails without needing to use the concept of probability. We who are less well informed about the toss's values for V and ω need to use the language of probability to make our prediction. This much is uncontroversial. The question is how a macro-probability, which does not conditionalize on point values for V and ω, can be objective.[39]

---

[37] A parallel claim is plausible concerning the objectivity of counterfactuals. Quine and others have noted that there are counterfactuals whose truth values are indeterminate (compare "If Verdi and Bizet were compatriots, they would be French" and "If Verdi and Bizet were compatriots, they would be Italian"). Some conclude from this that counterfactuals *never* describe objective matters of fact. I think this dismissal is too sweeping. For example, detailed counterfactuals that describe what would happen if you were to intervene in a causal system can be objectively true (Woodward 2003). For both conditional probabilities and counterfactuals, speakers often fail to be completely explicit, relying on shared information and context to supply relevant details.

[38] The Keller/Diaconis model does not assign probabilities to the initial conditions of a toss (its values for V and ω). It is not logically inevitable that the ratio of *areas* in the figure is the same as the ratio of *probabilities*.

[39] I treat the values of V and ω as "micro-descriptions" of a toss even though they apply to the whole coin toss apparatus without mentioning its parts. This is not objectionable since the problem concerning the reality of macro-probabilities arises from the fact that they are "coarse-grained" – i.e., their conditioning propositions are less contentful than (complete) micro-descriptions would be.

Figure 8: Newtonian coin tossing (from Diaconis 1998)
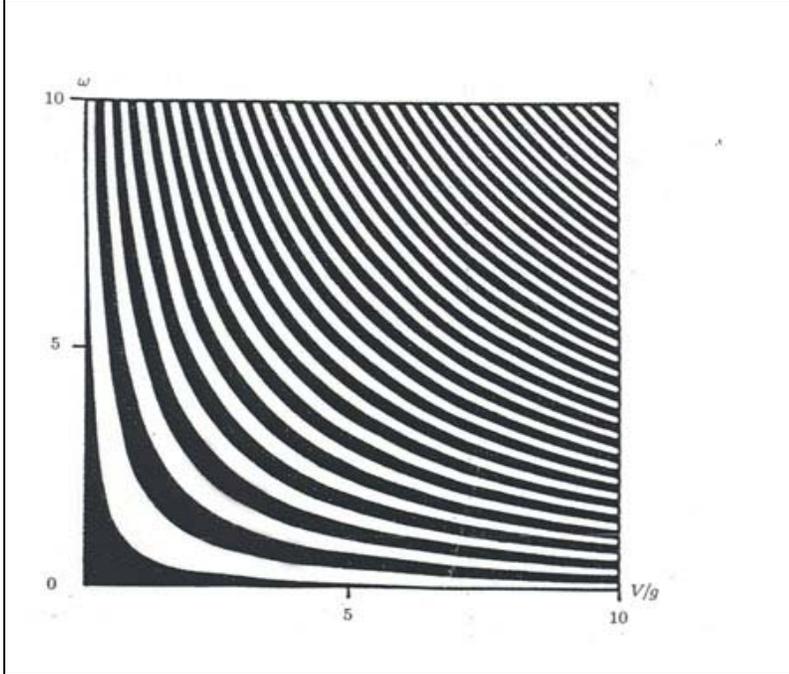
I want to argue, not just that the objectivity of macro-probabilities is *compatible* with Keller and Diaconis's deterministic model, but that this deterministic micro-model helps *justify* the claim that the macro- probabilities are objective. Consider the following representation of the relation between macro- and micro- dynamical theories.   Suppose there are n micro-states $(A_1 , A_2, \ldots, A_n)$ the system might be in at t_1. As before, X and Y are macro-properties.

(D) $Pr(Y \text{ at } t_2 \mid X \text{ at } t_1) = \sum_i Pr (Y \text{ at } t_2 \mid A_i \text{ at } t_1)Pr(A_i \text{ at } t_1 \mid X \text{ at } t_1).$

As noted earlier, the macro-probability $Pr(Y \text{ at } t_2 \mid X \text{ at } t_1)$ is a weighted average over various micro-probabilities of the form $Pr(Y \text{ at } t_2 \mid A_i \text{ at } t_1)$. The Laplacean grants that the first term on the right-hand side, $Pr(Y \text{ at } t_2 \mid A_i \text{ at } t_1)$, is objective. So the question of whether the macro-probability $Pr(Y \text{ at } t_2 \mid X \text{ at } t_1)$ is objective reduces to a question about the second term on the right – is the distribution of $Pr(A_i \text{ at } t_1 \mid X \text{ at } t_1)$ over $i = 1, 2, \ldots, n$ objective? That is, t*he diachronic macro-probability Pr(Y at t_2 ⎮X at t_1) is objective if a set of synchronic probabilities is too.*[40] This point holds regardless of whether the micro-theory is deterministic – i.e., independently of whether all the probabilities of the form $Pr(Y \text{ at } t_2 \mid A_i \text{ at } t_1)$ have extreme values.   Those with Laplacean sympathies may think that macro-probabilities can't be objective if the dynamic micro-laws are deterministic. But examining the decomposition (D) shows that

---

[40] Here I am relying on the assumption that if n quantities are each objective, so is any quantity that is a function of their values; "mind-independence" evidently has this compositional characteristic.

this reaction involves looking in the wrong place. Even if the micro-level dynamics are deterministic, what matters is the *initial conditions* -- the question is whether the distribution $Pr(A_i$ at $t_1 \mid X$ at $t_1)$ is itself objective.

If we apply the decomposition (D) to the case of coin tossing, and take account of the fact that V and ω are continuous quantities, we obtain the following double integral:

$$Pr(\text{heads at } t_2 \mid \text{tossed at } t_1) = \iint_{a,b} Pr(\text{heads at } t_2 \mid V{=}a \text{ and } \omega{=}b \text{ at } t_1)Pr(V{=}a \text{ and } \omega{=}b$$

$$\text{at } t_1 \mid \text{tossed at } t_1) \ (da)(db).$$

The question is whether the second product term on the right – the distribution of initial conditions – is objective. If it is, so is the macro-probability on the left. The Keller/Diaconis model helps simplify our question. Figure 8 shows that all we need to worry about is

$$Pr(\text{heads at } t_2 \mid \text{tossed at } t_1) = Pr(\text{heads at } t_2 \mid V \text{ and } \omega \text{ are in a black region at } t_1)Pr(V \text{ and } \omega \text{ are in a black region at } t_1 \mid \text{tossed at } t_1),$$

which simplifies to

$$Pr(\text{heads at } t_2 \mid \text{tossed at } t_1) = Pr(V \text{ and } \omega \text{ are in a black region at } t_1 \mid \text{tossed at } t_1).$$

Let me emphasize once more that the diachronic macro-probability on the left is objective if the synchronic probability on the right is.

Why should we think that $Pr(V$ and ω are in a black region at $t_1 \mid$ tossed at $t_1)$ is an objective quantity? Suppose that a large number of observers each toss the coin one or more times. The number of tosses may vary from observer to observer and the observers may obtain somewhat different frequencies of heads in their different runs. Suppose that experimenters who toss the coin 1000 times or more obtain frequencies of heads that are tightly clustered around 51%, while those who toss the coin a much smaller number of times obtain frequencies that are more widely dispersed. The similarity of the actual frequencies obtained by different observers should be explained by postulating a common cause. There is something about the process generating initial conditions on a toss that leads 51% of the tosses to be located in the black region of parameter space. This commonality is captured by the claim that $Pr(V$ and ω are in a black region at $t_1 \mid$ tossed at $t_1) = 0.51$. This probability claim applies not only to the runs of tosses that have been performed to date; it additionally makes a prediction about future runs of tosses, a prediction in which we are entitled to have some confidence.[41]

---

[41] I argued in Section 5 that our desire for general explanations should not be taken to show that macro-probabilities are objective. In the present section, I have argued that we can have evidence that a macro-probability is objective when it provides a common cause explanation of various observations. This may seem inconsistent (eschewing explanation in one place, espousing it in another), but I think it is not. Philosophers have examined the concept of explanation from two angles. The problem that Hempel and his successors addressed is to decide which of various true propositions belong in an explanation of some target proposition. The problem addressed under the heading of inference to the best explanation is to say which of various candidate hypotheses one should regard as true. I don't regard the requirement of generality as having an objective status in Hempel's problem, but I do think there is an objective reason, in many cases, to view unified explanations as better supported than disunified explanations. See Sober (2003b) for discussion.

The common cause argument just presented is epistemic – the point is to show that the kind of evidence we have concerning the objectivity of mass as a property also helps establish the objectivity of synchronic probabilities of the form $\Pr(A$ at $t_1 \mid X$ at $t_1)$. This point about evidence, however, does not answer a more metaphysical question. Where do macro-probabilities come from? What account can be given of how they arise?

Figure 9: How a synchronic macro-probability can arise from a micro common cause

---
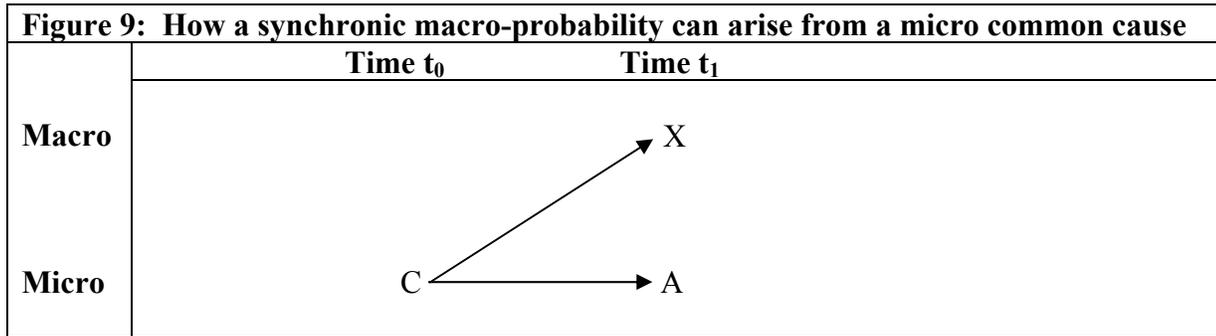
The idea of common causes provides an answer to this metaphysical question as well, one that is depicted in Figure 9. As before, let X be a macro-property at time t_1 and A a micro-property at that same time, but now consider C, which is a micro-variable at the earlier time t_0. C is a common cause of X and A, and C comes in n states ($C_1, C_2, \ldots, C_n$). We are interested in the question of how the macro-probability $\Pr(A$ at $t_1 \mid X$ at $t_1)$ arises. The answer[42] is that it arises from the common cause C:

$$\Pr(A \mid X) = \Pr(A\&X)\, /\, \Pr(X) = \sum_i \Pr(A\&X \mid C_i)\Pr(C_i)\, /\, \sum_i \Pr(X \mid C_i)\Pr(C_i).^{[43]}$$

Notice that all the probabilities on the right of the second equality sign are conditional micro-probabilities or unconditional probabilities of the micro-properties C_i. A Laplacean who believes in the reality of micro-probabilities can see from this how synchronic macro-probabilities of the form $\Pr(A$ at $t_1 \mid X$ at $t_1)$ arise. And once these are in place, the reality of diachronic macro-probabilities of the form $\Pr(Y$ at $t_2 \mid X$ at $t_1)$ is vouchsafed. We thus have in hand both a metaphysical and an epistemological answer to our question about the reality of macro-probabilities.

8. Concluding Comments

The Laplacean position (L_1 and L_2) allows that there is one circumstance in which the macro-probability $\Pr(Y$ at $t_2 \mid X$ at $t_1)$ represents an objective matter of fact. This occurs when the macro-probability $\Pr(Y$ at $t_2 \mid X$ at $t_1)$ and the micro-probability $\Pr(Y$ at $t_2 \mid A$ at $t_1)$ have the same value. If A at t_1 suffices for X at t_1, in accordance with the principle of mereological supervenience, this equality holds

---

[42] Or rather, this is *an* answer, one constructed to appeal to Laplaceans. I don't rule out the possibility that $\Pr(A$ at $t_1 \mid X$ at $t_1)$ has the value it does because of a *macro* common cause, thus providing an instance of "downward causation."
[43] Notice that the decomposition of $\Pr(A \mid X)$ described here does not require that the common cause render the two effects conditionally independent of each other. In fact, it will not, if each A_i either entails X or entails not-X. Whereas the epistemic argument is naturally understood in terms of effects being unconditionally dependent though conditionally independent, the metaphysical argument should not be understood in this way.

precisely when (Macro SO Micro) – that is, when $\Pr(Y$ at $t_2 \mid X$ at $t_1) = \Pr(Y$ at $t_2 \mid X$ at t_1 and A at $t_1)$. I have argued that this equality will rarely if ever be true. If we had true deterministic macro-theories, (Macro SO Micro) would automatically be correct. But we do not. And when both the macro- and the micro-theories are probabilistic, I know of no cases in which (Macro SO Micro) is correct; what is more, there seem to be very general reasons why this principle should fail. Facts about the micro-level seem constantly to provide predictively relevant information – information that is relevant above and beyond that provided by more coarse-grained macro-facts.

If (Macro SO Micro) were true, that principle would provide macro-probabilities and the theories in which they figure with a degree of autonomy. Predicting whether Y at t_2 would require just the macro-information that X at t_1; adding micro-details about what is true at t_1 would not be relevant. Anti-reductionists who yearn for this type of autonomy need to face up to the fact that it is not to be had. The autonomy of macro-level theories, if it exists, must be found elsewhere.

Laplaceans reason that if (Macro SO Micro) is false, then the macro-probability $\Pr(Y$ at $t_2 \mid X$ at $t_1)$ fails to describe anything objective. If we want to resist this conclusion, what are our options? One is to appeal to the usefulness of macro-probabilities in constructing explanations. I have offered two reasons for rejecting this suggestion. First, it is unclear why explanations constructed by using macro-probabilities are *objectively* better than explanations constructed on the basis of micro-probabilities. And second, it is a mistake to think that macro-probabilities are objective *because* they figure in our explanations. Rather, the situation is the reverse – if we want our explanations to be objectively correct, we can cite macro-probabilities only if we have some assurance that they are objective.

The line I prefer to take against the Laplacean focuses on the decomposition (D). If we assume, as the Laplacean does, that the micro-probabilities $\Pr(Y$ at $t_2 \mid A_i$ at $t_1)$ ($i = 1,2, \ldots, n$) are objective, it turns out that the macro-probability $\Pr(Y$ at $t_2 \mid X$ at $t_1)$ is objective precisely when the synchronic probabilities $\Pr(A_i$ at $t_1 \mid X$ at $t_1)$ ($i = 1,2, \ldots, n$) are too. Even if the micro-dynamical laws are deterministic, it is irrelevant to harp on that fact. The crux of the matter concerns the synchronic distribution of initial conditions – in Keller and Diaconis's coin tossing model, whether $\Pr(V$ and ω are in the black region at $t_1 \mid$ the coin is tossed at $t_1)$ is objective. I have argued that data can support the claim that such probabilities have objective values in just the way that data can support the claim that mass values are objective. What is more, the value of this synchronic macro-probability can be viewed as the upshot of an earlier micro-variable C. Laplaceans thus find themselves in an untenable position – if micro-probabilities are objective, so too must macro-probabilities be objective.

References

Albert, D. (2000): *Time and Chance.* Cambridge: Harvard University Press.

Brandon, R. (1990): *Organism and Environment.* Princeton: Princeton University Press.

Brandon, R. and Carson, S. (1996): "The Indeterministic Character of Evolutionary Theory -- No 'No Hidden Variables Proof' but no Room for Determinism Either." *Philosophy of Science* 63: 315-337.

Burnham, K. and Anderson, D. (1998): *Model Selection and Inference*. New York: Springer.

Butterfield, J. (2005): "Determinism and Indeterminism." *Routledge Encyclopedia of Philosophy*, forthcoming.

Crane, T. and Mellor, D. (1990): 'There is No Question of Physicalism." *Mind* 99: 185-206.

Darwin, C. (1859): *The Origin of Species*. Cambridge: Harvard University Press, 1959.

Dennett, D. (1987): "True Believers." In *The Intentional Stance.* Cambridge: MIT Press.

Diaconis, P. (1998): "A Place for Philosophy? The Rise of Modeling in Statistical Science." *Quarterly of Applied Mathematics* 56: 797-805.

Earman, J. (1986): *A Primer on Determinism*. Kluwer.

Earman, J. (2005): "Determinism – What We Have Learned and What We Still Don't Know." In J. Campbell, M. O'Rourke, and D. Shier (eds.), *Freedom and Determinism, Topics in Contemporary Philosophy Series* vol. II. Seven Springs Press. Also available at: www.ucl.ac.uk/~uctytho/detearmanintro.html.

Eells, E. (1981): "Objective Probability Theory Theory." *Synthese* 57: 387-444.

Forster, M. and Kryukov, A. (2003): "The Emergence of a Macro-World: A Study of Intertheory Relations in Classical and Quantum Mechanics," *Philosophy of Science* 70: 1039-1051.

Garfinkel, A. (1981): *Forms of Explanation.* New Haven: Yale University Press.

Giere, R. (1973): "Objective Single Case Probabilities and the Foundations of Statistics." In P. Suppes *et al.* eds., *Logic, Methodology, and Philosophy of Science* IV. Amsterdam: North Holland, pp. 467-483.

Glymour, B. (2001): "Selection, Indeterminism, and Evolutionary Theory." *Philosophy of Science* 68: 518-535.

Good, I. J. (1967): "On the Principle of Total Evidence." *British Journal for the Philosophy of Science* 17: 319-321.

Gould, S. (1980): *The Panda's Thumb.* New York: Norton.

Hacking, I. (////): "Do we see through the microscope?"

Hajek, A. (2003): "What Conditional Probability Could Not Be." *Synthese* 137: 273-323.

Horan, B. (1994): "The Statistical Character of Evolutionary Theory." *Philosophy of Science* 61: 76-95.

Jackson, F. and Pettit, P. (1992): "In Defense of Explanatory Ecumenism." *Economics and Philosophy* 8: 1-22.

Keller, J. (1986): "The Probability of Heads." *American Mathematical Monthly* 93: 191-197.

Kim, J. (1989), "The Myth of Nonreductive Materialism", *Proceedings and Addresses of the American Philosophical Association* 63: 31-47. Reprinted in *Supervenience and Mind*. Cambridge, England: Cambridge University Press, 1993.

Kim, S. (2000): "Supervenience and Causation -- a Probabilistic Approach." *Synthese* 122: 245-259.

Laplace, P. (1814): *A Philosophical Essay on Probabilities.* New York: Dover, 1951.

Levi, I. (1967): *Gambling with Truth.* New York: Knopf.

Levi, I. and Morgenbesser, S. (1964): "Belief and Disposition." *American Philosophical Quarterly* 1: 221-232.

Mayr, E. (1963): *Animal Species and Evolution.* Cambridge: Harvard University Press.

Pearl, J. (2000): *Causality – Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Putnam, H. (1975): "Philosophy and our Mental Life," in *Mind, Language, and Reality.* Cambridge, England: Cambridge University Press, 1975.

Putnam, H. (1971): *Philosophy of Logic.* New York: Harper.

Quine, W. (1953): "Two Dogmas of Empiricism." Cambridge: Harvard University Press, 20-47.

Reichenbach, H. (1956): *The Direction of Time.* Berkeley: University of California Press.

Rosenberg, A. (1994): *Instrumental Biology or the Disunity of Science.* Chicago: University of Chicago Press.

Rosenberg, A. (2001): "Indeterminism, Probability, and Randomness in Evolutionary Theory." *Philosophy of Science* 63: 536-544.

Salmon, W. (1984): *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton University Press.

Skyrms, B. (1980): *Causal Necessity.* New Haven: Yale University Press.

Sober, E. (1984): *The Nature of Selection.* Cambridge: MIT Press.

Sober, E. (1988): "The Principle of the Common Cause." In J. Fetzer (ed.) *Probability and Causation: Essays in Honor of Wesley Salmon*. Dordrecht: Reidel, 1988, 211-28; reprinted in *From a Biological Point of View*, Cambridge University Press, 1994.

Sober, E. (1992): "Screening-Off and Units of Selection." *Philosophy of Science* 59: 142-52.

Sober, E. (1993a): "Mathematics and Indispensability." *Philosophical Review* 102: 35-58.

Sober, E. (1993b): *Philosophy of Biology.* Boulder: Westview Press.

Sober, E. (1993c): "Temporally Oriented Laws," *Synthese* 94: 171-189; reprinted in *From a Biological Point of View*. Cambridge: Cambridge University Press, 1994, pp. 233-252.

Sober, E. (1999a): "The Multiple Realizability Argument against Reductionism." *Philosophy of Science* 66: 542-564.

Sober, E. (1999b): "Physicalism from a Probabilistic Point of View." *Philosophical Studies* 95: 135-174.

Sober, E. (2000): "Quine's Two Dogmas." *Proceedings of the Aristotelian Society* 74: 237-280.

Sober, E. (2001a): "The Two Faces of Fitness." In R. Singh, D. Paul, C. Krimbas, and J. Beatty (eds.), *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*. Cambridge University Press, vol. 2, pp. 309-321.

Sober, E. (2001b): "Venetian Sea Levels, British Bread Prices, and the Principle of the Common Cause." *British Journal for the Philosophy of Science* 52: 1-16.

Sober, E. (2003a): "Metaphysical and Epistemological Issues in Modern Darwinian Theory." In G. Raddick and M.J.S. Hodge (eds.), *The Cambridge Companion to Darwin*. Cambridge: Cambridge University Press, pp.267-287.

Sober, E. (2003b): "Two Uses of Unification." In F. Stadler (ed.), *The Vienna Circle and Logical Empiricism -- Vienna Circle Institute Yearbook 2002*. Kluwer, pp. 205-216.

Spirtes, P. Glymour, C., and Scheines, R. (2000): *Causation, Prediction, and Search.* New York: MIT Press.

Spirtes, P., and Scheines, R. (2003): "Causal Inference of Ambiguous Manipulations." *Proceeding of the Philosophy of Science Association*

Stamos, D. (2001): "Quantum Indeterminism and Evolutionary Theory." *Philosophy of Science* 68: 164-184.

Strevens, M. (2003): *Bigger than Chaos.* Cambridge: Harvard University Press.

Tsai, G. (2004): "Are Higher Level Explanations Always More General?" unpublished manuscript.

Wilson, R. (2004): *Boundaries of the Mind.* Cambridge: Cambridge University Press.

Woodward, J. (2003): *Making Things Happen.* Oxford: Oxford University Press.