

What Is Drift?

A Response to Millstein, Skipper, and Dietrich

Mohan Matthen[§]

The statistical interpretation of the Theory of Natural Selection claims that natural selection and drift are statistical features of mathematical aggregates of individual-level events. Natural selection and drift are not themselves causes. The statistical interpretation is motivated by a metaphysical conception of individual priority. Recently, Millstein, Skipper, and Dietrich (2009) have argued (a) that natural selection and drift are physical processes, and (b) that the statistical interpretation rests on a misconception of the role of mathematics in biology. Both theses are contested.

KEYWORDS

Drift • Natural selection • Statistical interpretation

1. The Constitution Thesis

Organisms are born; they die; they mate. These events change the make-up of organism-ensembles. Organisms enter an ensemble when they are born and exit when they die; by mating, organisms determine what kind of individual enters by a subsequent birth. This is how births, deaths and matings influence the make-up of organism-ensembles. Since evolution consists of changes to the make-up of organism-ensembles, let us call these events individual-level selection events or ILSEs. (The significance of the term 'selection' will become evident in a moment.) There are other kinds of ILSE – e.g., migration – but we need not worry about them here.

Organism-ensembles also change. For instance, suppose that we have an ensemble of 100 organisms, evenly divided into two kinds, K and K'. An organism of kind K (call it Kris) is killed by a predator; in the meanwhile, an organism of kind K' (call it Kristie) is born. If there are no other births or deaths, there are now 49 organisms of kind K, and 51 of kind K'. Kind K' has proportionately increased relative to kind K. This change of ensemble-wide proportions is an ensemble-level selection event or ELSE. That is, it is a change involving an ensemble of individual organisms.

What is the cause of this ELSE? Nothing other than that the predator killed Kris, and that Kristie became pregnant. That is: nothing other than ILSEs. After Kris's death and Kristie's birth – these events are both ILSEs – the ensemble-proportions change as a merely mathematical consequence. To put it in another way, the ensemble-level change is mere bookkeeping (cf. Walsh 2004); the causes of change are to be found at the individual level. It is because Kris's death and Kristie's birth result in changes in ensemble-level proportions that I call them individual-level selection events.

[§] Department of Philosophy, University of Toronto, 170 Saint George St., Toronto, ON M5R2M8, Canada

✉ E-mail: mohan.matthen@utoronto.ca

Received 3 January 2010; Accepted 22 March 2010



Let me make a metaphysical thesis of this:

Ensemble-level selection events are constituted without remainder by individual-level selection events; consequently, the causes of ELSEs are the causes of the ILSEs that constitute them. Thus, ELSEs are wholly caused by ILSEs.

Call this the Constitution Thesis. The Constitution Thesis is true because there is no need for an irreducibly ensemble-level cause to bring about frequency-changes. Indeed, it is difficult to see how there could be an irreducibly ensemble-level cause of frequency changes. A frequency change is, after all, wholly constituted by additions and deletions of individuals – there is no other way to bring it about. Additions and deletions of individuals have individual causes – a predator catches prey; two individuals mate. So the Constitution Thesis should not be construed as “an extreme form of nominalism” (as one referee disquietedly murmured). It is a simple consequence of the truism that frequency-changes are wholly constituted by additions and deletions of individuals.

The Constitution Thesis is not defeated by the following maneuver. Consider the following characterization of the Kris-Kristie scenario: more Ks than K’s were killed by predators; more K’-producing matings occurred than K-producing matings. This is an ensemble level characterization of the cause of the ELSE we are considering – namely that the proportion of K’s increased relative to Ks. Is there an ensemble-level event corresponding to this redescription, and if so does it defeat the claim that “ELSEs are wholly caused by ILSEs”? Not in my view. The ensemble-level event just described is wholly constituted by the individual-level events we have been considering.

Remember: ensembles are just pluralities. Suppose there are two pencils on my desk, and I push them both to the back of the desk. A lover-of-ensembles might insist that the plurality of pencils in my desk now has a different mean position. And he might insist that there is an ensemble-level cause of this ensemble-level event. My response to this challenge is to assent politely – but to add the codicil that the ensemble-level cause is wholly constituted by my pushes on the pencils.

It should therefore be noted that the Constitution Thesis is about ensembles (i.e., pluralities) not populations. Populations are causally connected networks of conspecific organisms; ensembles are merely collections of organisms. The claim that I want to defend below is that what evolutionary biologists call “random genetic drift” is a statistical feature of ensembles (see Matthen 2009 for a fuller argument). Drift can occur in an ensemble consisting of 1000 organisms that co-inhabit a locale and interact with each other – i.e., organisms who all belong to the same population. It can also occur in an ensemble consisting of 1000 organisms no two of which belong to the same species or reside in the same locale. It can even occur in an ensemble consisting of ensembles – there may be an expectation of how many of these ensembles go extinct, and this expectation may be defeated. Drift depends on the laws of mathematical statistics, not on causal interactions between organisms.

2. The Principle of Natural Selection

Some organisms possess heritable characteristics that make it easier for them to leave more descendants than others with different heritable characteristics. In ensembles, it is likely, but not certain, that the kinds better endowed for abundant reproduction will increase in numbers relative to those who are less so. (The term ‘better endowed for abundant reproduction’ is used here as a stand-in for the population-genetics term, ‘fitter’. I want to skirt the controversies that attend the use of the latter term, and I will not say more about what makes an organism or a type fit.)

Why? Because if it is probable (but not certain) that any organism of kind K will leave more descendants behind than one of kind K’, then it is probable (but not certain) that in a collection of organisms, the total number of descendants left by K-type individuals will exceed the number of those left by K’-type individuals. (This is a consequence of what one might call the probability-frequency rule: more probable events probably occur more frequently than less probable events.) This will lead to a greater proportion of K-type individuals. This simple mathematical truth is the Principle of Natural Selection.

According to what I shall call the Crude Statistical Interpretation of the Principle of Natural Selection, the causes of the increase of better endowed kinds are none other than the causes of births, deaths, and matings of individual organisms. The Crude Statistical Interpretation relies on the Constitution Principle above. The increase of K-type individuals relative to K'-type individuals is an ELSE: that is, it is a fact or event concerning an ensemble. By the Constitution Thesis, it is wholly caused by ILSEs. In other words, the increase in K-type individuals is wholly caused by the individual births, deaths, and matings that occur within the relevant ensemble.

The statistical interpretation does not reify natural selection. Natural selection is not a cause of these ensemble-level changes. Indeed, it is not a process or event or force – it is not the kind of thing that could be a cause. ILSEs cause ensemble-level change. The Theory of Natural Selection is a theory about expectations regarding how the course of individual-level events will influence ensemble-level proportions. The Theory does not posit causes of its own.

What then is drift? I said earlier that when one kind is better endowed than another for abundant reproduction, that kind is likely, but not certain, to increase proportionately to the other. Statistically speaking, one might expect the better endowed kind to increase, but such expectations can be (and often are) violated in actual fact. Departures from expected values are what population geneticists call 'drift'. (This use of the term overlooks an important distinction due to Roberta Millstein, and I will refine my terminology in the next section.) What is it that causes kind K' to increase relative to kind K despite the latter being better endowed? On this point too, the statistical interpretation stands firm. The Constitution Thesis forbids a search for ensemble-level causes of ensemble-level changes of frequency – at least of ensemble-level causes that are not wholly constituted by ILSEs. The causes are all at the individual level. In particular, it is a mistake to look for a process that acts on ensembles, processes entitled 'drift' or 'natural selection'.

Now, the Crude Statistical Interpretation stands in need of some nuance. In particular, the statement that “the causes are all at the individual level” needs qualification (cf. Matthen and Ariew 2009). Moreover, there are some population-level effects in natural selection. But leave this aside. The broad-brush characterization given above is good enough for the purposes of this article. All that is important here is the statement that terminates the preceding paragraph, namely that 'drift' and 'natural selection' are not names of ensemble-level causes. All versions of the statistical interpretation, whether crude or refined, hold this.

3. The Process Interpretation

In a recent paper, Roberta Millstein, Robert Skipper, and Michael Dietrich (2009) – henceforth MSD09 – argue that random genetic drift is “a physical process where heritable physical differences between entities are causally irrelevant to differences in reproductive success” (p. 2; page numbers refer to the pdf version of the article). This goes against the statistical interpretation. In the statistical interpretation, drift is a departure from expected values attributable to the uncertainty inherent in the expectations outlined above. This uncertainty is inherent in any series of births, deaths, and matings. It is not a separate “physical process” over and above the individual-level causes of such events. As a proponent of the statistical interpretation, I shall query the philosophical motivation of the process view articulated in MSD09.

Roberta Millstein (2002) makes an important distinction between “drift as outcome” and “drift as process”. Drift-as-outcome is a departure from the expected results of selection. Drift-as-process is the cause of such departures. But what is the cause of departures from expected values? Such a departure might be something like the following. Suppose we have a population consisting of half Ks and half K's. Suppose further that K and K' are equally fit. We expect that at some future time, the population will consist of 50% Ks and 50% K's. Say that in fact it consists of exactly 40Ks and 60K's at this later time. Then we could say that the K-drift (as outcome) is -10, and K'-drift is +10. (We needn't worry here about normalizing this to the ensemble.) What is the cause of this? This cause is, according to Millstein, drift-as-process. Note that, according to the crude statistical interpretation, the underlying ILSEs are the whole cause: that is to say, events such as the death of Kris and the birth of Kristie. Note, therefore, that if the drift-as-process posit is to have any bite against the crude statistical interpretation, it has to be accompanied by the thesis that drift-as-process is not wholly constituted by ILSEs of the sort just mentioned.

MSD09 cites three quotations from biology text-books to set up their process interpretation.

◆ Douglas Futuyma: “The genes included in any generation, whether in newly formed zygotes or in offspring that survive to reproduce, are a sample of the genes carried by the previous generation. Any sample is subject to random variation, or sampling error” (p. 2; all quotations are taken from MSD09; I have not verified them at their source). According to the authors of MSD09, “sampling error” here is drift-as-outcome. Then they remark: “Futuyma invokes indiscriminate parent sampling by giving an example where changes in gene frequency in a population of snails are the result of being squished by cows, a process in which the color of the snails is causally irrelevant.” The idea seems to be this: since any type of snail is just as likely as any other to be squished by a cow, one would expect all snail-types to be equally affected by the cows. In other words, cow-squishing is a “indiscriminate physical process”. Now, it may turn out that some types of snail are, by chance, more affected by cow-trampling than others. This distorts the effects of selection.

◆ Continuing: “(Joan) Roughgarden similarly explains indiscriminate gamete sampling in a finite population before asserting, “Genetic drift is the name for changes in gene frequency caused by this sampling error” (p. 2). The point of this quote is difficult to understand. Earlier, “sampling error” was (correctly) identified as drift-as-outcome. Here, it is given as the cause thereof. Nonetheless, note that Roughgarden speaks of causes, but we are not told clearly what she takes the causes of drift-as-outcome to be.

◆ Finally, a quote from Mark Ridley: “When selection is acting at a locus, random sampling also influences the change in gene frequencies between generations” (p. 2). Here, “random sampling” seems to be the cause of drift-as-outcome.

MSD09 summarize: “for all three of these biologists, drift is indeed partially characterized as an outcome (a change in gene frequencies), but it is an outcome caused by a certain type of physical process (indiscriminate parent or gamete sampling) – not as outcome alone” (p. 3; italics in the original). (Actually, this is false: Roughgarden confusedly posits sampling error as a cause; Ridley seems to come close to the same error, but his use of the term ‘random’ is amplified by the authors into a full-blown theory of drift-as-process.)

The question is this. What is there in any of the above quotes to suggest that random sampling is a physical process, or any kind of process, for that matter? The only physical events cited here are cow-tramplings, and these are indeed indiscriminate. It seems that MSD09 identifies cow-trampling with drift-as-process in this particular case. But they have no support from their texts, at least as far as these texts are quoted.

4. Two Views of Drift

Taking Millstein’s distinction between drift-as-outcome and drift-as-process on board, here’s the picture that seems to emerge from MSD09. Suppose that there is selection in favor of A-colored snails as against B-colored snails because the A-type is better camouflaged against predator-birds. According to MSD09, this would be a discriminating physical process: it affects different types differently. (Some philosophers take such processes to be selection, as Walsh 2004, p. 352, and footnote 15, shows.) In this ensemble, there are also various other “processes” at work – cow-trampling is one of them. Now, let’s suppose that this is an infinite ensemble. Then indiscriminate processes such as cow-trampling would affect snail-types exactly equally. (This follows directly from definition of ‘indiscriminate’.) Accordingly, cow-trampling would not distort the outcome of selection. In finite ensembles, by contrast, the expected outcome might not actually occur because, by chance, a larger proportion of A-type snails may happen to be knocked out by trampling cows. Thus, cow-trampling brings about the unexpected outcome – i.e., it brings about drift-as-outcome. So, cow-trampling is drift-as-process. The position just outlined is the process interpretation. This process interpretation is, for three reasons, metaphysically misguided.

First, it is unclear why indiscriminate physical processes should be invoked to explain unexpected results. For every cause of death, whether discriminate or indiscriminate, there is an expected result. In the case of poor-camouflage predation, the expected result is a decline of B-colored snails. In the case of cow-

trampling, the expected result is an equal effect on snails regardless of color. The process interpretation cites indiscriminate processes as the root cause of unexpected results in selection. But these indiscriminate processes can do this work only if they themselves culminate in unexpected values – for they interfere with selection only when, contrary to expectation, they affect different types unequally. The question is: what accounts for unexpected results in cow-trampling and other such indiscriminate processes? The process interpretation has no answer.

Notice that the statistical interpretation does not face this difficulty. It holds that there can be uncertainty in the results of any process, discriminate or indiscriminate. (I won't address here the question of where this uncertainty comes from. See, however, Matthen 2009.)

Second, consider exactly the same physical factors acting in finite and infinite ensembles. In infinite ensembles, as noted before, there would be no drift-as-outcome – the results would conform to expectations. But there would (by the stipulation that the same factors are at work) nevertheless be cow-trampling. So there is drift-as-process in the infinite ensemble – remember, cow-trampling is drift-as-process in the process interpretation – but no drift-as-outcome. Drift-as-process is at work in all ensembles, according to this view, but it produces drift-as-outcome only in finite ensembles. This implies that indiscriminate sampling cannot explain drift-as-outcome by itself. Ensemble size is needed as well. But once ensemble size is invoked, the indiscriminate/discriminate distinction becomes irrelevant. Both kinds of process depart from the expected in exactly the same fashion for exactly the same (statistical) reasons.

Once again, note that the statistical interpretation does not face this difficulty. Mathematical statistics proves that the smaller a population, the greater the probability of an unexpected result. And this is true whether a process is discriminate or indiscriminate.

Finally, why is something like cow-trampling a process? I understand a process to be the propagation of a single causal influence. You switch the heat on under a kettle and after some time, the water in the kettle boils. This is a process because it is the propagation of heat from the burner. Let a ball free on an inclined plane and it rolls to the bottom. This is a process because the action of gravity on the ball propels it down its path. In cases like this, unified causal influences play themselves out on some object or objects. Cow-tramples, by contrast, are disconnected events. Daisy the cow grazes on her patch of grass, trampling on some snails as she does so; Betsy, another cow, grazes on a different patch of grass, trampling other snails as she goes. Daisy does not influence Betsy; Betsy does not influence Daisy. This is like two pots of water on two burners – they both come to the boil, but there is no one process. In the same way: why would Daisy-Betsy be considered a single process?

The statistical interpretation, of course, aggregates these events: it counts up the deaths of snails and the numerical consequences thereof. But it has no reason to unify these events. According to the statistical interpretation, these aggregates consist of events, many of which are independent of others. Such collections of events are merely aggregates – “heaps” of events collected together for some extrinsic reason (such as co-location) or no reason at all. (Refined statistical interpretations, such as that offered in Matthen and Ariew 2002, 2009, do allow for some ensemble-level causes in a derivative way – not natural selection or drift though.)

5. Conclusion

MSD09 entitle the statistical interpretation the “Drift as Outcome Alone” view, which they archly abbreviate “DOA”. (For readers fortunate enough to be unfamiliar with North American hospital jargon, “DOA” stands for “Dead On Arrival”.) This would have been more accurately put in this way: no cause of drift except those wholly constituted by ILSEs. Of course, NCODETWCBILSE is not as wittily derisive as DOA.

And they say: “Since DOA cannot be justified by an appeal to common conceptions among biologists and philosophers, the most charitable interpretation is that DOA is in accord with the mathematical models of drift” (p. 4). According to MSD09, the statistical interpretation is unconcerned with the physical reality of natural selection and drift. MSD09 makes much of the contention that “the biologists who developed the mathematical models of drift did so with the intention of modeling physical processes (the indiscriminate

sampling processes) that they took to be occurring in nature” (p. 7, italics in the original). Accordingly, the statistical interpretation is criticized as follows: “it is a mistake to think that we can glean definitions of drift from mathematics alone” (p. 5).

All of this is highly misleading. The statistical interpretation fully recognizes that a mathematical-statistics model represents biological reality, but does not constitute it; it knows that the assumptions that lie behind a statistical model must reflect the biological reality that is being modeled or represented. It is hard to see how anybody could think otherwise. The statistical interpretation is driven, at least in its crude form, by the Constitution Thesis, and similar metaphysical assumptions, as well as by various theses concerning causation that I have not discussed here. (See, however, Matthen and Ariew 2002, 2009.) There is no need at all for a “charitable interpretation”: what is needed is an argument against the Constitution Thesis or against this application, or against the notions of causation employed by proponents of the statistical interpretation. Millstein (2006) appears perfectly well aware of this; so it is hard to see why this “charitable interpretation” scat is being sung here.

However that may be, Shapiro and Sober (2007) argue against the Crude Statistical Interpretation’s use of the Constitution Thesis, and Millstein (2006) offers some reflections on natural selection and causation – both more cogent than MSD09. Matthen and Ariew (2009) contains rejoinders to both.

The statistical interpretation is concerned with physical reality: it simply holds that natural selection and drift are best understood in terms of the statistical properties of mathematical aggregations of individual-level selection events. Physical reality is to be found in the causes of these ILSEs. In some refined statistical interpretations, causes (but derivative, not physical, causes) are to be found also at the ensemble level. But natural selection and drift are not processes in any statistical interpretation.

Literature cited

- Matthen, M. 2009. Drift and “statistically abstractive explanation”. *Philosophy of Science* 76: 464-487.
- Matthen, M. and A. Ariew 2002. Two ways of thinking about fitness and natural selection. *Journal of Philosophy* 49: 55-83.
- Matthen, M. and A. Ariew 2009. Selection and causation. *Philosophy of Science* 76: 201-224.
- Millstein, R. L. 2002. Are random drift and natural selection conceptually distinct? *Biology and Philosophy* 17: 33-53.
- Millstein, R. L. 2006. Natural selection as a population-level causal process. *British Journal for the Philosophy of Science* 57: 627-53.
- Millstein, R. L., R. A. Skipper Jr., and M. R. Dietrich 2009. [\(Mis\)interpreting Mathematical Models: Drift as a Physical Process](#). *Philosophy and Theory in Biology* 1:e002.
- Shapiro, L. and E. Sober 2007. Epiphenomenalism: The dos and the don’ts in thinking about causes. In: *From Greek Philosophy to Modern Physics* Ed. by G. Wolters and P. Machamer. University of Pittsburgh Press.
- Walsh, D. 2004. Bookkeeping or metaphysics? The units of selection debate. *Synthese* 138: 337-361.

Copyright © 2010 Author(s).

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs license, which permits anyone to download, copy, distribute, or display the full text without asking for permission, provided that the creator(s) are given full credit, no derivative works are created, and the work is not used for commercial purposes.

ISSN 1949-0739

