Why *you* have one trait rather than another: The failure of the explanatory-chain strategy

Casey Helgeson[a,b]

unpublished manuscript, November, 2012

ABSTRACT. In *The Nature of Selection* (1984), Sober argued that natural selection is in principle powerless to explain why any individual organism has the traits it does rather than the very same individual having different traits. A debate ensued, in which critics have argued against Sober by laying explanations end-to-end, to form a chain of explanation that begins with selection and passes through one or more intermediate events before reaching the target explanandum. I argue that Sober's critics misunderstand how contrastive explananda (why p rather than q) behave in such explanatory chains, and that this strategy has so far failed.

1 Introduction

In *The Nature of Selection* (1984), Sober contrasts two strategies for explaining why a roomful of schoolchildren all read at the third-grade level. To give a *developmental* explanation, you would attend to one child at a time and describe the earlier events that led to that child's current reading proficiency. To give a *selectional* explanation, you would note that reading at the third-grade level was a prerequisite for admittance to the class—children at other reading levels were excluded. But we shouldn't think that the two strategies address the same question:

> The developmental story says why each individual has one reading level rather than another. The selectional story, on the other hand, shows why the room is filled with individuals reading at the third grade level rather than with different people with different reading abilities. (Sober 1984, 150)

Selective admittance explains why Sally is in the classroom rather than having been turned away, but it does not explain why she reads at one level rather than another. Similarly

---

[a]Department of Philosophy, University of Wisconsin–Madison
[b]Author's note: Manuscript reformatted and archived in March 2020. There is a bit of overlap between this manuscript and the published articles: Helgeson (2013, 2015). While the main topic here—the explanatory-chain structure of arguments for the positive view—is not addressed by those articles, the upshot of this manuscript is similar to that of Helgeson (2015).

(reasons Sober), natural selection in biological evolution does not explain why any particular individual has the traits it does (say, a short, thick beak, and a large body) rather than the very same individual having different traits (say, a long, narrow beak, and a smaller body). Call this claim **S** (for Sober). What natural selection does explain, say Sober and subsequent advocates of **S**, is why this individual before us with the short beak and large body is present in the population rather than some other individual with a long beak and a smaller body. More generally, what selection can explain is why a population consists of one set of individuals (with such and such features) rather than another set (with different features).

There is an active opposition to **S**, and the *manner* in which critics argue against **S** has two conspicuous features. First, Sober's critics argue almost exclusively via what I'll call the *explanatory-chain strategy*. They accept that Sober's reasoning makes direct appeal to natural selection unconvincing, but they say that selection *does* explain some intermediate explanandum (what the intermediate is varies from argument to argument), and then they argue that this intermediate in turn explains Sober's explanandum. Since the explanation relation is *transitive*, Sober's critics continue, selection really does explain Sober's explanandum, so **S** is false.[1] Second, critics give short shrift to Sober's "rather than" clause. Sober's explanandum is *contrastive* (why p rather than q). But critics chop off the contrastive clause and address their arguments to the leftover stump "why p," sometimes adding—as though an afterthought—that by "why p," they mean "why p rather than q."[2]

This abbreviation of the explanandum would be unproblematic if arguments about "why p" were to generalize straightforwardly to "why p rather than q." I argue that they do not. Unless Sober's critics are simply talking past him, they must squarely address the contrastive explanandum. Once they do this, I argue, the explanatory chain arguments fall apart. In what follows, I show that *when viewed contrastively*, the explanatory chain arguments are invalid. These arguments rest (implicitly) on certain trans-world identity assignments, but offer no theory of identity that could underwrite those assignments. This is particularly bad news for opponents of **S**, not only because there is no theory on the table that would do the job, but because independence from such metaphysical assumptions is one of the selling points of their position (Forber 2005).

2 Preliminaries

Sober's classroom illustration motivates **S**, but there is also a more substantive argument, a quick rehearsal of which will supply the background needed for my analysis of the

---

[1]Neander (1988, 1995b,a), Matthen (1999), Forber (2005), and Nanay (2005, 2010) all pursue variations on this strategy; one exception is Birch (2012).

[2]Neander (1988) presents Sober's position and argues against it all without mention of contrastive explananda. Nanay (2005) does the same. Neander (1995a) acknowledges Sober's contrastive clause, yet frames her argument against **S** in non-contrastive terms. Nanay tacks the contrastive clause onto the end of his previous argument in a follow-up note (Nanay 2010). Forber (2005) presents Sober's position without mention of the contrastive clause, but later alludes to it in some parenthetical comments (pp.336–339). The exceptions to the "short shrift" generalization are Matthen (1999) and Birch (2012).

explanatory-chain strategy used against **S**. Let genetic trait t be an adaptation, and organism O an individual that has t.[3] Does selection explain why O has t rather than another trait? Consider a counterfactual history in which an alternative selection regime shaped the population. Having another selection regime in place just means that different individuals will survive and reproduce. This different selection regime can change O's genotype only (barring horizontal gene transfer) by changing who survives and reproduces *among O's ancestors*. But if any of O's ancestors fail to survive and reproduce, then O will never come to be. Thus an alternative selection regime cannot lead to O having different traits, and therefore the actual selection regime does not explain why O has t rather than another trait (Sober 1995).

Notice that the argument depends on the counterfactual scenario lacking any organism that can count as *the same as* (numerically identical with) organism O. Though Sober doesn't endorse any particular theory of trans-world identity for individual organisms, the following constraint on trans-world identity assignments would suffice for generating the required non-identity rulings: looking across alternative possible scenarios, "two" organisms can count as the same individual only if they have the same parents (Matthen 1999). Call this constraint *ancestor essentialism*.[4]

Regarding the "explains" relation, the argument supposes that among the necessary conditions for "w explains why p rather than q" is that had w failed to obtain, then q would have occurred instead of p. I too will operate with this understanding of contrastive explanation, but I want to acknowledge that it is unclear exactly what sort of counterfactual is required, and to suitably relax the requirement. It may be good enough that had w' obtained (for some particular w' mutually exclusive from w), then q would have occurred. And perhaps it is enough that were w' to occur, then q would be *probable*, or at least *more probable* than otherwise. I assume only that some such counterfactual is a necessary condition for "w explains why p rather than q," and I introduce a placeholder for the appropriate type of counterfactual conditional. So that we don't forget this is just a placeholder, I use the non-standard 'umbrella handle' arrow ($\hookrightarrow$) to denote the relation. Nothing in what follows will depend on the precise nature of the $\hookrightarrow$ relation.[5]

3   The explanatory-chain strategy

The literature includes several instantiations of the explanatory-chain strategy, but they all admit of the same analysis. So I will discuss only the original version—what I'll call the *Origin Argument* (Neander 1988, 1995b,a; Forber 2005). That argument goes as follows. Again we have organism O with trait t. Stipulate that the actual past selection regime explains the first appearance of t in some past generation, and that O inherits t from members of that past

---

[3]I limit the discussion to genetic traits in order to bracket complications introduced by development.

[4]Matthen (1999) calls it *origin essentialism*, but there are a number of different theses with that name, and in the present case it is more accurate to say *ancestor essentialism*, so I switch to that term.

[5]Also note that in the literature arguing about **S**—as in the broader literature on explanation—"explains" means *helps explain*, or *partly explains*; it does *not* mean explains fully. I follow this usage.

generation. The argument then goes:

(1) Selection explains the presence of trait t in some past generation.
(2) The presence of t in that past generation explains (via heredity) why O bears t.

(3) Selection explains why O bears t.

The general form of the argument is:

(1) w explains why x
(2) x explains why p

(3) w explains why p.

Supposing the "explains" relation is transitive, the argument form is valid.[6] But this can't be the whole argument. Unless critics are simply talking past Sober, the argument above must be read as a gesture at some other, nearby argument that employs contrastive explananda. A fully fleshed-out Origin Argument must have the form:

(1) w explains why x rather than z.
(2) x explains why p rather than q.

(3) w explains why p rather than q.

This form, however, is *invalid*. To illustrate, I repurpose the setup of the well-known trolley problem (Foot 1967).

A runaway trolley (streetcar) is barreling down an unused section of track towards five oblivious workers. You stand at a lever that would divert the trolley to another track, where there is just one oblivious worker. The only way you can influence the outcome is by pulling the lever. Say you pull the lever and the lone worker is killed. The next day, there is a funeral. So the *actual* events were: your pulling the lever, the death of the worker, and the funeral. Bringing in contrastive clauses, let the final explanandum be why there was a funeral *rather than no funeral* (why p rather than q). This is explained by the death of the worker (x explains why p rather than q). Had *no one died*, there would have been no funeral. And let the intermediate explanandum be why one worker was killed *rather than five* (why x rather than z). This is explained by your pulling the lever (w explains why x rather than z). Had you stood by, then five would have died. So both premises are true. Yet the conclusion is false: your pulling the lever does not explain why there was one funeral rather than no funeral. Had you not pulled the lever, there would have been *five funerals*.

The counterexample works by employing two different alternatives to the actual intermediate event x (the death of the one worker). One of those alternatives (five workers die) is relevant only to the first link in the explanation chain, while the other (no deaths) is relevant only to the second link.[7] (See Figure 1 for a visual aid.) This allows for there to be an appropriate

---

[6]I won't press the point, but the transitivity of explanation is controversial (*e.g.*, Putnam 1973).

[7]The insight behind this counterexample comes from recent work on the metaphysics of causation (see Hall 2000; Maslen 2004; Schaffer 2005) in which alleged counterexamples to the transitivity of causation (A causes

4

true counterfactual associated with each link in the chain, yet no such counterfactual that runs the whole length of the chain. In other words, *don't pull ↪ five deaths* and *no deaths ↪ no funeral* are both true, while *don't pull ↪ no funeral* is not.)

pull lever ———————→ **one death** ———————→ **one funeral**
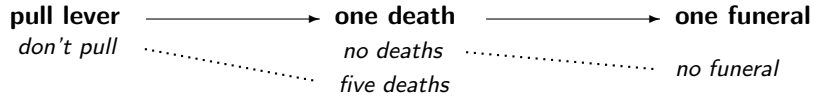*don't pull* ·············· *no deaths* ···················· *no funeral*
five deaths

Figure 1: Schematic illustration of a counterexample to the validity of the explanation-chain argument form. (Dotted lines are counterfactuals.)

Notice that if both links in the chain were to use the same alternative to the actual intermediate event, then we could avoid the kind of disconnect just described. With this in mind, consider the following quick fix. Both alternatives to the actual intermediate event satisfy the coarse-grained description "lone worker survives." So why not lump them together under that heading and consider them one and the same? This seems to make the problem go away: now both links in the chain use the same alternative intermediate event (see Figure 2).
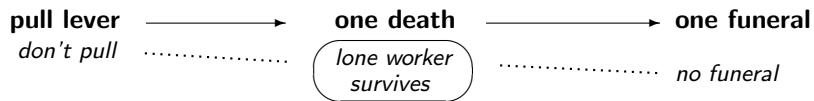
pull lever ———————→ **one death** ———————→ **one funeral**
don't pull ·············· (*lone worker survives*) ···················· *no funeral*

Figure 2: The trolley example again, now using the coarse-grained description "lone worker survives" to encompass both alternatives to the actual intermediate event.

The trouble with this fix is that it introduces an equivocation. There are relevantly different ways in which the lone worker might have survived, and not pulling the lever leads to a very specific one (in which five other workers die), which is incompatible with any of the ways of instantiating *the lone worker survives* that might have led to a final outcome of *no funerals*. So this fix does not work. It does, however, camouflage the problem. Imagine that the beginning-to-end counterfactual *don't pull ↪ no funerals* were not transparently false. And say we believe the ↪ relation is transitive. In this imaginary situation, we might well look at Figure 2 and find ourselves thinking: "It appears I've shown that *don't pull ↪ no funerals* must be *true*—and that pulling the lever *does* explain why there was one funeral rather than no funerals." I suggest that Sober's critics have fallen victim to a mirage like this one.

To see how, I must first reformulate the Origin Argument using contrastive explananda. The contrastive clause in Sober's explanandum is *rather than the same organism having another trait*. To keep things simple, I'll treat having t as one trait, and *lacking t* as the alternative. Here is my first-pass reformulation:

B, and B causes C, but A doesn't cause C) are disarmed by arguing that the two instances of 'B' are really not the same relatum. If the two instances of 'B' are really different relata, then what we have is A causes B, and C causes D, in which case it is no challenge to transitivity that A doesn't cause D. These authors disambiguate the two instances of 'B' by introducing contrastive clauses—one instance of 'B' is really 'B rather than B′′' while the other is really 'B rather than B⋆′' (where B′ and B⋆ are different counterfactual events).

(1) Selection explains why trait t was present in the past population rather than absent.
(2) The presence of t in the past generation explains (via heredity) why O has t rather than O lacking t.

(3) Selection explains why O has t rather than O lacking t.

Given what I've assumed about the explanation relation, the argument's conclusion is true only if, for some relevant alternative to the actual selection regime, the counterfactual *different selection regime $\hookrightarrow O$ lacks t* is true. The counterfactuals associated with the first and second premises are: $(1')$ *different selection regime $\hookrightarrow$ past population lacks t*, and $(2')$ *past population lacks $t \hookrightarrow O$ lacks t*. Verbally, the consequent of $(1')$ and antecedent of $(2')$ are the same. But in light of the preceding discussion, we should ask whether this broad event description conceals an equivocation (see Figure 3).

actual regime ⟶ t **present** ⟶ O **has** t
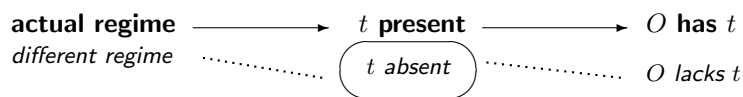*different regime* ············· (t *absent*) ············· O *lacks* t

Figure 3: Schematic illustration of (a contrastive reconstruction of) the Origin Argument. Dotted lines are counterfactuals. Does the event description t *absent* hide an equivocation?

Whether the Origin Argument equivocates turns on trans-world identity assignments for individual organisms. To see this, suppose—for just a moment—ancestor essentialism. Starting with the first link in the chain, a different selection regime cannot (assuming ancestor essentialism) deliver a t-less past population without also precluding the existence of every individual that bears t in the actual past population (see §2). But some of those past t-bearers are ancestors of O. In other words, that first link can be true only if the intermediate counterfactual population lacks some of O's ancestors. But, of course, only a past population that *includes* all of O's ancestors could lead to a final outcome in which O is present but lacks t. So the second link can be true only if the intermediate counterfactual population includes all of those ancestors. Consequently, there is no way to instantiate the event description "t absent" (Figure 3) that will render both links in the chain true. The Origin Argument equivocates (supposing ancestor essentialism).

The Origin Argument succeeds only if it does not equivocate, and to show that it does not equivocate is to show that the kind of t-less past population that can result from an alternative selection regime is also one that can lead to a final outcome that includes organism O. Doing that requires saying what it takes to count as organism O. The Origin Argument involves a sleight of hand that obscure this requirement.

There is currently no theory of identity on the table that would do the job. Matthen's (1999) argument against ancestor essentialism offers only a hint of a suggestion for an alternative theory that may be more favorable to opponents of **S**.[8] No other opponent of **S** has engaged the topic. Yet without a theory of trans-world identity, the Origin Argument is invalid.

---

[8]See Lewens (2001) and Pust (2004) for attempts to flesh out, then criticize, Matthen's hint.

4   Conclusion

Sober's critics argue against his position primarily via the explanatory-chain strategy (the Origin Argument is a representative instance). I attribute the intuitive plausibility of these chain arguments to their non-contrastive formulation: When the links in the chain are formulated non-contrastively as "w explains x" and "x explains p," and when it is added only as an afterthought that by "p" one means "p rather than q," then the "rather than" clause of the *intermediate* explanandum gets swept under the rug. This intermediate contrastive clause is where the equivocations hide.

The Origin Argument is invalid until the equivocation problem is fixed. The same goes for other instances of the explanatory-chain strategy (Matthen 1999; Nanay 2005, 2010). If my analysis is correct, opponents of **S** would seem to have three options: (a) give up the claim to do without metaphysical assumptions, then devise and defend a theory of trans-world identity that will underwrite the identity assignments required by their arguments, (b) abandon the explanatory-chain strategy and go back to the drawing board, or (c) stick to non-contrastive explanatory chains and concede that their arguments talk past Sober's position.

References

Birch, J. (2012). The negative view of natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences 43*(2), 569–573.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*.

Forber, P. (2005). On the explanatory roles of natural selection. *Biology and Philosophy 20*(2), 329–342.

Hall, N. (2000). Causation and the price of transitivity. *The Journal of philosophy 97*(4), 198–222.

Helgeson, C. (2013). What selection can and cannot explain: A reply to Nanay's critique of Sober. *Philosophy of Science 80*(1), 155–159.

Helgeson, C. (2015). There is no asymmetry of identity assumptions in the debate over selection and individuals. *Philosophy of Science 82*(1), 21–31.

Lewens, T. (2001). Sex and selection: A reply to Matthen. *The British Journal for the Philosophy of Science 52*(3), 589–598.

Maslen, C. (2004). Causes, contrasts, and the nontransitivity of causation. In J. Collins, N. Hall, and L. A. Paul (Eds.), *Causation and counterfactuals*, pp. 341–58. MIT Press: Cambridge, Mass.

Matthen, M. (1999). Evolution, Wisconsin style: Selection and the explanation of individual traits. *The British Journal for the Philosophy of Science 50*(1), 143–150.

Nanay, B. (2005). Can cumulative selection explain adaptation? *Philosophy of Science 72*(5), 1099–1112.

Nanay, B. (2010). Natural selection and the limited nature of environmental resources. *Studies in History and Philosophy of Biological and Biomedical Sciences 41*, 418–419.

Neander, K. (1988). What does natural selection explain? Correction to Sober. *Philosophy of Science 55*(3), 422–426.

Neander, K. (1995a). Explaining complex adaptations: A reply to Sober's 'reply to Neander'. *The British Journal for the Philosophy of Science 46*(4), 583–587.

Neander, K. (1995b). Pruning the tree of life. *The British Journal for the Philosophy of Science 46*(1), 59–80.

Pust, J. (2004). Natural selection and the traits of individual organisms. *Biology and Philosophy 19*(5), 765–779.

Putnam, H. (1973). Reductionism and the nature of psychology. *Cognition 2*(1), 131–146.

Schaffer, J. (2005). Contrastive causation. *The Philosophical Review 114*(3), 327–358.

Sober, E. (1984). *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. University of Chicago Press.

Sober, E. (1995). Natural selection and distributive explanation: A reply to Neander. *The British Journal for the Philosophy of Science 46*(3), 384–397.