

Status Quo Bias, Rationality, and Conservatism about Value*

Jacob M. Nebel

Many economists and philosophers assume that status quo bias is necessarily irrational. I argue that, in some cases, status quo bias is fully rational. I discuss the rationality of status quo bias on both subjective and objective theories of the rationality of preferences. I argue that subjective theories cannot plausibly condemn this bias as irrational. I then discuss one kind of objective theory, which holds that a conservative bias toward existing things of value is rational. This account can fruitfully explain some compelling aspects of common sense morality, and it may justify status quo bias.

In 1985, the Coca-Cola Company replaced its classic formula with New Coke. The new soft drink was sweeter than the original, and it had performed much better in blind taste tests. But this change was a great failure in marketing strategy. American consumers reacted negatively not because they disliked the taste of New Coke, but rather because they opposed the withdrawal of the old formula (which was later reintroduced—quite successfully—as Coca-Cola Classic). The Coca-Cola Company had failed to account for status quo bias.

There has been much empirical research on the descriptive explanation for status quo bias. Some, for example, believe that status quo bias is explained by our strong aversion to losses (compared to our desire for gains), by our tendency to believe that harmful actions are worse than harmful inactions, or by other cognitive biases. I discuss some of these questions below, but my main interest is in normative questions related to status quo bias. Does status quo bias compromise the ratio-

* I am greatly indebted to Derek Parfit and Larry Temkin, each of whose work inspired much of my thinking here, and whose generous comments were extremely helpful. I am also grateful for comments from Boris Kment, Michael Otsuka, Bastian Stern, Trevor Teitel, and the reviewers and editors at *Ethics*. I owe special thanks to Thomas Kelly for encouragement and invaluable feedback on several drafts.

nality of the preferences that it tends to support? In other words, are our resulting preferences less than fully rational? What nonnormative features distinguish status quo bias from an uncontroversially rational preference for the current state of affairs? Should the evidence for status quo bias decrease our confidence in certain views in normative ethics? These questions are normative because they are about what we have reason to choose, to prefer, and to believe.

I argue that, in some cases, status quo bias is fully rational. I begin, in Section I, by trying to clarify the concept of status quo bias. In Section II, I argue that subjective accounts of reasons cannot plausibly condemn this bias as irrational. This is an interesting result, given (a) the importance of subjective accounts of rational preference in economics and (b) economists' tendency to regard status quo bias as irrational. In Section III, I discuss one kind of objective theory, which holds that a conservative bias toward existing things of value is rational. I argue that this theory supports status quo bias in some cases. It also has important implications for commonsense morality. In Section IV, I discuss whether it can be rational to change our preferences depending on which outcome is the status quo. This discussion bears on the distinction between criticizable "sour grapes" preferences and cases of healthy preference adaptation. It also has implications for practical reasoning on certain theories of value.

I. UNDERSTANDING STATUS QUO BIAS

Status quo bias is a disposition, or tendency, to prefer some state of affairs because it is the status quo. This definition has several components, which I try to clarify below.

The status quo is the current, existing state of affairs. Different outcomes, however, may count as the status quo with respect to different alternatives or criteria. For example, a Coca-Cola drinker's preference may track the status quo with respect to the drink's branding more than it tracks the status quo with respect to the drink's taste. This kind of case may raise important questions about what follows, as in other contexts.¹ We

1. Here are some examples of similar questions in other contexts. In Kantian ethics, how should we determine the morally relevant descriptions of our actions? In statistics, when determining the probability of someone having a heart attack within the next ten years, which traits should determine the comparison class whose heart attack frequencies are relevant? In safety-theoretic epistemology, when are two belief-forming methods sufficiently similar, so that false beliefs produced by one method imply that true beliefs produced by the other are nonetheless in danger of error? The importance of such questions cautions against the idea that we should not care about which state of affairs is the status quo on the grounds that different states may be regarded as the status quo according to different criteria.

can set such questions aside for now. In most of the cases I discuss, there is a clear status quo ante. When it is ambiguous, we can discuss whether some relevant criterion picks out one outcome as the status quo for purposes of the comparison in question. Or, alternatively, we can treat our bias in favor of the status quo as a bias in favor of what we regard as the status quo: the description under which we think of some outcome may motivate our preference for it.

Many experiments have confirmed that we have a tendency to prefer the status quo, at least in certain kinds of situations.² The experiments follow this general pattern:

In a *neutral framing*, some subjects are given a set of options, with no option designated as the status quo. In a *status quo framing*, other subjects are given the same set of options, but one option is designated as the status quo. The probability that some option is selected is significantly higher when it is designated as the status quo than in a neutral framing or when some other option is designated as the status quo. When some different option is designated as the status quo for other subjects, its popularity increases significantly.³

Social scientists have observed this effect not only in the marketing of soft drinks but also in financial investments, urban planning, electric power service, tort law reforms, and university health plans. These experiments show a pattern in our decision making, or what we tend to choose or to do. This pattern may be evidence of a more fundamental pattern in our preferences.

When we prefer some state of affairs because it is the status quo, the sense of “because” has to do with our motives. It implies that our preference is (at least partly) caused or explained by the fact or belief that this state is the status quo. I shall not assume that the agent is always aware of what motivates her preference. Nor shall I assume that when we prefer the status quo because it is the status quo, we prefer it only because it is the status quo: there may be other beliefs that play a role in the explanation of our preference. But, in order to manifest status quo bias, the contents of these beliefs must necessarily depend on or be sufficiently similar to the proposition that the outcome is the status quo. We may have good evidence that a preference manifests status quo bias if the following counterfactual is true: if we learned that this outcome was not the status quo, then we would not have this preference. By contrast, we

2. See Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler, “Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias,” *Journal of Economic Perspectives* 5 (1991): 193–206.

3. William Samuelson and Richard Zeckhauser, “Status Quo Bias in Decision Making,” *Journal of Risk and Uncertainty* 1 (1988): 7–59.

often prefer the status quo because we believe it has other features that make it, all things considered, better than its alternatives, and we would prefer this option even if it were not the status quo. This counterfactual is good evidence that the preference is not an instance of status quo bias.

Some might opt for a narrower conception of status quo bias.⁴ It may seem that the phenomenon studied by economists and psychologists, which many of them regard as irrational, is a disposition to prefer the status quo merely or simply because it is the status quo. On this view, a preference manifests status quo bias only if the belief that the preferred outcome is the status quo is the sole motivation for the preference. But I doubt that the critics of status quo bias have this narrower conception in mind, for three reasons. First, this interpretation is inconsistent with psychologists' and economists' attempts to explain status quo bias in terms of other (seemingly irrational) considerations that motivate us to prefer the status quo. I shall mention these considerations at the end of this section. Second, the empirical evidence, which is largely based on observed discrepancies in choice behavior between neutral framings and status quo framings, does not seem to support the claim that *this* bias, so narrowly understood, is widespread. It is not as if, in a status quo framing, we have in mind only one feature of the outcome we choose—namely, that it is the status quo. It is more plausible that our preferences are motivated by a variety of facts that have some close explanatory connection to this feature. Third, when psychologists and economists argue that status quo bias is irrational, their arguments extend beyond this narrower conception.⁵ Their normative arguments are not sensitive to the distinction in question. So it would be surprising if their beliefs about the irrationality of status quo bias were restricted in this way. For these reasons, I do not restrict my claims about status quo bias to cases in which the mere fact that some outcome is the status quo is the sole motivation for the agent's preference.⁶

My topic, however, is normativity, not motivation. The explanatory or causal role of the status quo does not entail that the fact that some outcome is the status quo is a reason—that is, that it really counts in favor of the preference—or even that we believe that this fact is a reason. But this explanatory connection between the preference and the status

4. Thanks to Bastian Stern and Trevor Teitel for pressing this worry.

5. In addition to the arguments I discuss in Sec. II and Sec. IV, see Ilana Ritov and Jonathan Baron, "Protected Values and Omission Bias as Deontological Judgments," *Psychology of Learning and Motivation* 50 (2009): 133–67.

6. Nonetheless, my arguments might imply that even the narrower disposition is, in some cases, fully rational. This depends on whether it can be rational to prefer some outcome even if what directly motivates the preference is not itself a good-making feature of the outcome. I am inclined to think that it can. But I shall not here try to resolve this question, or other interesting questions about the narrower conception of status quo bias.

quo is, I believe, relevant to the rationality of status quo bias. Even if we prefer some outcome that happens to be good, this preference seems irrational if our preference has no close connection to the facts which make it good, or our reasons for preferring it, and is instead motivated by some normatively irrelevant fact. Opponents of status quo bias believe that preferences manifesting this bias are not responsive to the normatively relevant facts.

I shall argue that status quo bias is, in some cases, fully rational. But my arguments do not show that the mere fact that some outcome is the status quo is itself a reason to prefer this outcome. Rather, in some cases of status quo bias, there is a deeper reason that justifies the preference, which I shall explain in Section III. It may nonetheless be true that we have this preference because the preferred outcome is the status quo, in the sense of “because” described above. Just as other beliefs may play a role in explaining the motivation for preferences that manifest status quo bias, other reasons may play a role in justifying preferences that manifest status quo bias. The reason I shall give in Section III necessarily depends on the fact that the outcome is the status quo: necessarily, if this outcome were not the status quo, then this deeper reason would not exist. Some may prefer to say that such cases do not manifest status quo bias because a bias is, by definition, irrational. My aim is not to analyze what it is to be a bias, so I will simply make a terminological stipulation: I shall use the word “bias” in a way that does not trivially entail irrationality, and no inherent rationality is built into my definition of status quo bias.

One way to see whether some argument supports the rationality of status quo bias is to subject them to Nick Bostrom and Toby Ord’s proposed test for irrational status quo bias:

The Reversal Test. When we prefer the status quo to some change in a certain parameter, consider a change to the same parameter in the opposite direction. If the status quo is still preferred, then we must explain why our position cannot be improved through changes to this parameter. If we are unable to do so, then we have reason to suspect that we suffer from an irrational bias for the status quo.⁷

For example, Bostrom and Ord discuss resistance to technologies that would increase human intelligence. They ask critics of cognitive enhancement whether they would support a proposal to decrease human intelligence. Since most of them would oppose that change, they believe that

7. Nick Bostrom and Toby Ord’s version, in “The Reversal Test: Eliminating Status Quo Bias in Applied Ethics,” *Ethics* 116 (2006): 656–79, 665, is specific to evaluative beliefs about outcomes. I have generalized the test to apply to preferences, however understood.

the status quo is a local optimum. According to Bostrom and Ord, these critics then have the burden to justify this belief, and failure to do so is evidence that their opposition to cognitive enhancement is due to irrational status quo bias. I shall explain, roughly, what counts as a failure below.

Importantly, the Reversal Test is not intended to be a definition of status quo bias. It is also not a necessary or sufficient condition for status quo bias. Rather, it is intended to give us evidence of irrational status quo bias. It is supposed to be a good (but fallible) guide, or way of telling, whether or not our preferences are irrational. Bostrom and Ord argue that the Reversal Test is a good guide for detecting irrational status quo bias because the status quo (in terms of some parameter) is *prima facie* unlikely to be the optimal state of affairs. So the onus is on those who prefer the status quo to justify this preference. If they cannot do so, then we may reasonably conclude that their preference for the status quo is irrational. They are disposed to prefer the status quo; they would not prefer this outcome if it were not the status quo; and their preference is not sensitive to any deeper reason to prefer the status quo. In other words, they suffer from irrational status quo bias.

But what kind of explanation or justification is sufficient to pass the Reversal Test? Here are a few considerations that often count in favor of maintaining the status quo:

- Transition costs. It might be better if the United States used the metric system. But switching to the metric system would have massive transition costs, which might outweigh the benefits of the metric system.
- Uncertainty. I have only owned Apple computers. I don't know what it would be like to own a PC—it might be better in some ways, worse in other ways. But, since I am more than satisfied with Apple, I prefer this safe option.
- Cognitive limitations and deliberation costs. Whenever I cycle from my residence to any particular locale, I prefer to take the same route. This path may not be the shortest or most pleasant route, but it would be difficult for me to deliberate about it every time I need to go somewhere. So I prefer to maintain the status quo because it is easier than spending the mental energy on this decision.

When some preference does not pass the Reversal Test, we might reasonably suspect that our preference is guided by irrelevant factors, not by good reasons (like those above). Critics of status quo bias seem to assume that the following kinds of considerations are among such irrelevant factors—and, therefore, provide evidence of irrational status quo bias:

- Existence and longevity. People tend to regard existing things as good and to regard things that have existed longer as better.⁸ But the fact that something exists, or has existed for a long time, does not seem to make the status quo good, or better than the alternatives.
- Sunk costs. The fact that we have invested resources in some project (many philosophers, psychologists, and economists assume) is not a reason to continue with that project—that is, to maintain the status quo with respect to that project.⁹
- Loss aversion. We tend to weigh losses more heavily than their corresponding gains. When a departure from the status quo has expected gains that compensate for the expected losses, then (Bostrom and Ord assume) it is at least as good as the status quo.
- Omission bias. We tend to weigh the harms of omissions as less severe than the harms of actions, but that is a mistake. We can often maintain the status quo by doing nothing, but (many argue) we should not prefer the status quo on that basis.¹⁰
- Status quo. We may suffer from some brute bias that favors the status quo, which is not grounded in one of the factors above. But the fact that some state of affairs currently obtains does not seem to be a reason to prefer that state of affairs.

Many people believe that these considerations are not good-making features of outcomes. When we use these features to justify our preferences after applying the Reversal Test, we exhibit irrational status quo bias. I note these considerations because they will be helpful when arguing that some theory supports status quo bias in certain cases.

To summarize this section, status quo bias is a disposition to prefer some outcome because it is the status quo. This does not imply that we prefer the outcome merely because it is the status quo. Our preference may be sensitive to and justified by some deeper reason which, necessarily, only obtains when the outcome is the status quo. We can use the Reversal Test to identify preferences that are not responsive to any such deeper justification. Failing the Reversal Test is not a necessary or sufficient condition for status quo bias, but it provides evidence of irrational status quo bias. This is because a successful justification for the preference is sufficient to pass the test.

8. See Scott Eidelman and Christian S. Crandall, "Bias in Favor of the Status Quo: Bias and the Status Quo," *Social and Personality Psychology Compass* 6 (2012): 270–81.

9. For a dissenting view, see Thomas Kelly, "Sunk Costs, Rationality, and Acting for the Sake of the Past," *Noûs* 38 (2004): 60–85.

10. See Ritov and Baron, "Protected Values."

II. SUBJECTIVE THEORIES

Following Derek Parfit, we can distinguish between two kinds of views about the rationality of preferences.¹¹ Objective theories hold that our reasons to prefer some outcome to some alternative are ultimately grounded in the features of these outcomes. According to subjective theories, however, we have no reasons for our preferences, except in a derivative way when our having some preference would help us fulfill some other preference.

Subjectivists deny that our preferences can be intrinsically irrational—that is, irrational simply in virtue of the things we prefer. Hume famously held that I can rationally prefer “the destruction of the whole world to the scratching of my finger,” or “my own acknowledged lesser good to my greater.”¹² But most subjectivists would deny that “anything goes” with respect to the rationality of our preferences, because they think there are certain constraints that rational preferences must satisfy. In this section, I consider two of those constraints, which prohibit uninformed and arbitrary preferences. I argue that neither of them can plausibly rule out status quo bias.

A. *Idealization*

Subjectivists may appeal not to our actual preferences, but rather to the preferences we would have in certain ideal conditions. These ideal conditions may include full information, clear thinking, and perfect instrumental rationality. If these conditions serve as rational constraints on our preferences, then perhaps status quo bias is irrational because we would not have this bias under such ideal conditions.

Full information may help in some cases of status quo bias. It seems that we sometimes prefer the status quo because we know what it is like, whereas the alternatives are uncertain. This preference may partly explain, for example, why one might hold onto a low-paying job rather than searching for a new one. But that is not the whole explanation, which would require some discussion of transition costs and other features. And information limitations may not play an important role in many cases of status quo bias. Economists have found evidence of status

11. See Derek Parfit, *On What Matters* (Oxford: Oxford University Press, 2011), vol. 1. Some people think that the rationality of preferences comes apart from what we have most reason to prefer. These people could reinterpret my claims in terms of what we ought to, have most reason to, or correctly or reasonably prefer. On the distinction between rationality and reasonableness, see W. M. Sibley, “The Rational versus the Reasonable,” *Philosophical Review* 62 (1953): 554–60.

12. Book 3, pt. 3, sec. 3 of Hume’s *A Treatise of Human Nature*. But see Parfit, *On What Matters*, vol. 2, sec. 111, for evidence that Hume believed that we have some object-given reasons to have our desires.

quo bias even in cases where there seem to be no relevant asymmetries in empirical information that, if corrected, would change the agent's preference.¹³ It is not as if the agent cares about some nonnormative property *F*, knows whether the status quo is *F* and does not know whether the alternative is *F*. So, in many cases of status quo bias, we do not have sufficient evidence to conclude that the agent would not prefer the status quo if she were better informed. Full information, therefore, does not seem enough to condemn status quo bias as irrational.

Subjectivists might concede that full information is not sufficient to prevent status quo bias. But they might hope to eliminate status quo bias by fiat. Subjectivists might simply enumerate the things that we must idealize away from, and include on the list the status quo, or our knowledge of what the status quo is. This strategy, however, would yield a trivial answer to our question about the rationality of status quo bias. If the idealization process eliminates our knowledge of the status quo, then of course the preferences resulting from status quo bias will not survive idealization. This is relevantly like idealization in Rawls's original position: we idealize away from morally irrelevant factors, such as our physical or mental natural endowments in the status quo. As Rawls recognizes, however, it is a substantive normative claim that these factors are morally irrelevant.¹⁴ Similarly, the subjectivist's requirement that we idealize away from the status quo assumes, rather than justifies, a substantive claim about rationality.

Furthermore, this strategy would not give us a satisfying story for what makes status quo bias irrational. When subjectivists appeal to our idealized preferences, they typically claim that a particular pattern of concern is irrational in virtue of some more general feature of the idealization process. If I knew the relevant facts and were fully instrumentally rational, for example, I might not desire to drink as much coffee as I do. The irrationality of my desire to drink lots of coffee, on this view, is grounded in—and, hence, explained by—a more general principle. On the simple enumeration strategy we are considering, however, the irrationality of status quo bias is not grounded in a general principle that criticizes some feature of this bias: it just belongs on the list, by the subjectivist's fiat. That answer is not an explanation. When I ask why my desire to drink lots of coffee is irrational, it would be unsatisfying to answer that our idealized selves are, by stipulation, not coffee drinkers.

I conclude, then, that idealization is not a plausible way for subjectivists to earn their right to criticize status quo bias as irrational.

13. Samuelson and Zeckhauser, "Status Quo Bias," 35.

14. John Rawls, *A Theory of Justice*, rev. ed. (Cambridge, MA: Belknap Press of Harvard University Press, 1999), 86. Thanks to an anonymous reviewer for this analogy.

B. Arbitrariness

Subjectivists might argue that status quo bias is irrational when and because the fact that some state of affairs currently obtains is arbitrary. For example, Samuelson and Zeckhauser find that when choosing the color of a new car, we tend to prefer a color that is arbitrarily chosen as the status quo. And one might think, following Michael Smith, that it is irrational to have preferences that draw arbitrary distinctions.¹⁵ Fabienne Peter, for example, holds that status quo bias is irrational because it is arbitrary.¹⁶

To test this argument, consider another case where the charge of arbitrariness arises. One putative counterexample to subjectivist theories is the following case due to Parfit:

Future Tuesday Indifference. Suppose I want to avoid pain on any day, except on any future Tuesday, in which case I am indifferent. I would prefer several hours of agony next Tuesday to five minutes of slight pain experienced any other day of the week. I know what pain is like, and I know that the calendar might easily have been different, so that some other day would've been Tuesday. I have no false non-normative beliefs about Tuesdays (e.g., that a capricious deity will punish those who try avoid suffering on Tuesdays). But I am indifferent towards my suffering on future Tuesdays.¹⁷

Subjectivists might agree that Future Tuesday Indifference is irrational.¹⁸ But, they might say, the irrationality lies not in any object-given reason to avoid pain for its own sake, but rather in the arbitrariness of wanting to avoid pain except on any future Tuesday.

Can subjectivists, however, help themselves to a value-free conception of arbitrariness? According to Parfit, for any possible objects of preference x and y , it is arbitrary to prefer x to y just in case there are no facts about x and y that count in favor of this preference.¹⁹ Since these facts, on Parfit's view, must be object-given reasons, subjectivists cannot appeal to arbitrariness (so understood). But perhaps subjectivists have in mind a different conception of arbitrariness, which does not require any object-given reasons.

15. See Michael Smith, "Desires, Values, Reasons, and the Dualism of Practical Reason," *Ratio* 22 (2009): 98–125.

16. See Fabienne Peter, *Democratic Legitimacy* (London: Routledge, 2005), 109.

17. Derek Parfit, *Reasons and Persons* (Oxford: Clarendon, 1986), 124.

18. For a dissenting view, see Sharon Street, "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters," *Philosophical Issues* 19 (2009): 273–98.

19. Parfit, *On What Matters*, 1:79.

This value-free conception of arbitrariness might have to do with metaphysical naturalness, as pioneered by David Lewis.²⁰ Suppose that a property is arbitrary to the extent that it is unnatural, or fails to carve reality at its joints. Being *grue*, for example, is unnatural. Something is *grue* just in case it is either (a) green and observed before 3000 AD or (b) blue and observed after 3000 AD. It seems arbitrary and, therefore, irrational to want to collect *grue* emeralds. Similarly, the property *being on a Tuesday* is unnatural, so preferences that distinguish on the basis of that property are arbitrary and, therefore, irrational.

As an analysis of the relevant sort of arbitrariness, this view fails. Consider

Negative Nancy, whose greatest aim is to count negatively charged particles. Negative Nancy doesn't have any false nonnormative beliefs about negative charge. She doesn't find this activity particularly soothing or pleasurable,²¹ but that is how she wants to spend her time.

Assuming that negative charge is a perfectly (or close to perfectly) natural property, Negative Nancy's preference passes the test. But it seems to me that Negative Nancy's preference is—like Future Tuesday Indifference and counting *grue* emeralds—arbitrary and, therefore, irrational.²²

The subjectivist could instead claim that unnaturalness is a sufficient, but not necessary, condition for arbitrariness. Negative Nancy's preference is arbitrary even though it does not track a gruesome distinction. We should then ask whether status quo bias tracks a gruesome distinction, because that would be sufficient for arbitrariness.

20. See David Lewis, "New Work for a Theory of Universals," *Australasian Journal of Philosophy* 61 (1983): 343–77.

21. Following Parfit, *On What Matters*, vol. 1, sec. 6, I distinguish between desires, which (according to subjectivists) give us reasons, and the likings or dislikings of sensations that make things pleasant or painful. We have no object-given reasons to like or dislike such sensations (e.g., to be calmed by the sound of rain, or to like the bitterness of beer), but we have object-given reasons to want (not) to experience sensations that we (dis)like. The subjective theories under consideration would not find it intrinsically irrational to want to experience a sensation that we find painful rather than pleasant.

22. An anonymous reviewer suggests that some subjectivists would deny that these preferences are arbitrary and, therefore, irrational. But whatever the subjectivist says on behalf of *grue* emeralds, negatively charged particles, and future Tuesdays could presumably apply to the status quo. For example, an agent with status quo bias might be soothed by the mere fact that some option maintains the status quo, or she might find it important to maintain the status quo (whatever it is) because of some tradition started by her ancestors. In opting for more liberal constraints on rationality, the subjectivist permits status quo bias.

I don't think it does. When our preferences are affected by status quo bias, we prefer some outcome because it currently obtains (although not necessarily merely because it currently obtains, for the reasons given in Sec. I). But *currently obtaining* seems much more natural than, say, *being grue*. We can even suppose that time's passage is an illusion. On this view, the present time is not metaphysically privileged; "now" is just like "here," in that both refer to what they do only because of the time or place of the thought or utterance.²³ On the view that time's passage is an illusion, the concept *now* (like the concept *here*) would have no application if there were no thinkers or utterers. When we manifest status quo bias, we prefer some state of affairs because it obtains at the time of the preference. But that preference still does not seem particularly gruesome. It is no more gruesome than self-regarding preferences, because "I" is also relative to some thinker's thoughts. Ruling out status quo bias on such grounds would, therefore, rule out self-regarding preferences in its wake. But it seems perfectly rational to prefer some state of affairs on the grounds that it is better for me.²⁴

It does not seem that status quo bias is arbitrary in virtue of drawing unnatural distinctions. But it may be arbitrary for some other reason. Return to the example of car color: people prefer to maintain the arbitrarily chosen status quo with respect to car color. In some cases, the existence of a state of affairs is itself arbitrary. This arbitrariness is common for habits, customs, and other traditions. In a minority of countries, for example, people drive on the left side of the road. The choice between right- and left-hand traffic was, we may suppose, arbitrary in the following sense: the initial legislators of traffic laws were not motivated by any belief about the benefits of one side over the other; in their eyes, it may as well have been decided by flipping a coin.

Now, there may be good reasons for these countries to maintain left-hand traffic—mostly about the transition costs of a shift to right-hand traffic. But consider

Traffic Genie. Some genie offers the UK Parliament an opportunity to switch to right-hand traffic without incurring any transition costs. No one would mistakenly drive on the left, except people who would have mistakenly driven on the right in the status quo; traffic signs would be magically reversed; and drivers would wake up with vivid memories of driving on the right.

23. See Parfit, *Reasons and Persons*, 178.

24. A different argument for the view that status quo bias rests on a gruesome distinction might appeal to the fact (mentioned at the beginning of Sec. I) that different states of affairs may be regarded as the status quo according to different criteria. But I have already given, in n. 1, other examples in which matters of great importance, which do not seem gruesome, are sensitive to contextually relevant criteria in similar ways.

Should Parliament prefer left-hand traffic because it is the status quo? Many would say No. The fact that some state is the status quo does not seem to be a reason to prefer that state of affairs; it doesn't seem to count in favor of that state's continuation. On the subjectivist view we are considering, that is because the initial choice of traffic laws was arbitrary in the sense that it was not motivated by any belief about the benefits of one side over the other.

This lack of a motivating belief in the past, however, does not explain the lack of a reason today. First, even if there had been some motive for the initial choice of traffic laws, that might not make the preference today any more rational or less arbitrary. For example, the motive might have been some irrational superstition held by left-handed people, which no one currently believes. This fact would not make the current preference rational or nonarbitrary, even if some closely related fact (e.g., about tradition) might. Second, it would commit the genetic fallacy to infer that there is no normative reason today simply because there was no initial motivating belief.²⁵ We should not accept a conclusion about the rationality of some practice today based merely on some fact about the origin of that practice.

The kind of arbitrariness that compromises rationality is not the lack of a motive or explanation for a preference. In the cases we are considering, what makes status quo bias seem arbitrary (and, therefore, seem irrational) is the lack of an object-given reason to prefer the status quo for its own sake. Subjectivists cannot appeal to such reasons, so they cannot plausibly claim that status quo bias is irrational because it is arbitrary.

III. CONSERVATISM AND STATUS QUO BIAS

It may seem obvious that status quo bias will be irrational on an objective theory of reasons. The fact that some state of affairs currently obtains is not an object-given reason to prefer that state of affairs. Put simply, existence does not make an outcome good. In this section, however, I argue that our disposition to prefer the status quo may be fully rational on a plausible objective theory.

G. A. Cohen defends the following conservative principle:²⁶

Conservatism: We have reason to preserve valuable things that currently exist, even when we could replace them with things of equal or greater value.

25. Thanks to an anonymous reviewer for this suggestion.

26. G. A. Cohen, "Rescuing Conservatism: A Defense of Existing Value (All Souls Version)," in *Finding Oneself in the Other*, ed. Michael Otsuka (Princeton, NJ: Princeton University Press, 2012).

According to Conservatism, things that are valuable—in a wide sense of “things,” which includes practices and traditions—warrant a kind of response other than promoting or producing more of them.²⁷ They warrant preservation, conservation, loyalty, respect, or cherishing.²⁸ And these responses ought to be directed toward the valuable things as the particular things they are, not as mere carriers of value. If we regard valuable things as fungible, as if they can be exchanged for substitutes of equal or greater value, then we treat them as mere carriers of value; we care only about the value, not about the thing that bears it. We have a reason, for example, to prevent the destruction of great works of art (e.g., the *Mona Lisa*), persons (e.g., Jane Doe from Boise), cultures (e.g., that of the Byzantine Empire), or institutions (e.g., the University of Oxford) even if equally good or slightly better things of their kind would thereby replace them. This reason is not absolute, and conservatives may disagree about the strength of the reason—in general and in specific contexts. But the distinctive claim in Conservatism is that there is such a reason, which is not given by our subjective desires and which systematically favors the status quo.

Conservatism does not mention reasons to prefer states of affairs, but it is plausible that if Conservatism is true, then we have reason to prefer states of affairs that involve the preservation of valuable things. (This may be an instance of a more general principle, which I discuss below, that we have reason to prefer states of affairs in which people do what they have reason to do.) Conservatism supports our disposition to prefer the current state of affairs, because we are not rationally required to be indifferent between the status quo and a replacement of equal value. According to Conservatism, we have reason to prefer the pres-

27. Cohen restricts his Conservatism to things that are intrinsically valuable—i.e., valuable in virtue of their intrinsic features. Erich Hatala Matthes, in “History, Value, and Irreplaceability,” *Ethics* 124 (2013): 35–64, has argued that this restriction is both over- and underinclusive. We could avoid Matthes’s objections by restricting Conservatism instead to things that have final value—i.e., value for their own sakes. But I wish to remain neutral on this question, so I simply use “valuable” to stand in for the relevant kind of value, whatever it is.

28. We might wonder whether this is plausible in the context of Coca-Cola, the example with which I began this article. I think it is not, because the original Coke is not valuable for its own sake. But some might view the original Coke as a kind of cultural artifact or part of our tradition. On this view, the Coca-Cola Company may have mistakenly seen the soft drink merely as a means to the satisfaction of our taste buds, when it was really a valuable aspect of our culture. This might explain the otherwise surprising virulence of the backlash to New Coke. Alternatively, as Michael Otsuka has suggested to me, Cohen’s conservative account of personal value might explain this case. Either way, it would not be an objection to my view if it does not support status quo bias in the Coca-Cola example. My aim is to defend the rationality of status quo bias in some cases, not in all or most of them. Thanks to Thomas Kelly for raising this question.

ervation of existing things even when departing from the status quo—for example, by creating some new thing—would result in things of equal or greater value. Sometimes the status quo does not preserve valuable things: in cases of injustice or harm, it preserves bearers of disvalue. And sometimes maintaining the status quo would destroy something of value, which we can only prevent by departing from the status quo. Perhaps that kind of case is an instance where different outcomes can be described as the status quo, depending on the relevant features or alternatives: we might describe the prevention of the thing's destruction as the status quo, since the relevant criterion is the continued existence of the valuable thing. But even if the unambiguous status quo option involves the destruction of valuable things, this merely implies that Conservatism does not support every status quo preference; it may nonetheless support the disposition or bias for a certain class of cases.

Here are two examples of the phenomenon I have in mind.²⁹ First, in the Book of Job, God allows Satan to destroy Job's family and wealth. But, in the end, God gives Job twice as many animals as he had before, along with ten children. Should Job prefer the resulting outcome? In one important respect, No. Once you have a family, it doesn't make the outcome better if your children are killed and replaced with more children. Job had reason to prefer the preservation of his family and to regret its replacement. Second, William Butler Yeats's poem "For Anne Gregory" contains the lines, "Love you for yourself alone / And not your yellow hair." When you love a particular individual, you don't just love him or her for the intrinsic properties he or she possesses. You would not be better off, with respect to love, if the person you love were replaced by someone whose properties were in some way better and in every other way at least as good. As Cohen suggests, valuing something may, in this respect, be like loving someone.³⁰

To see why Conservatism supports status quo bias, consider how it applies to the Reversal Test. We imagine a change to some parameter in both directions—for example, an increase and a decrease in human intelligence. Some opponent of the increase in human intelligence also holds that a decrease would be bad. Bostrom and Ord are suspicious: it is *prima facie* unlikely, they reason, that the current level of human intelligence is as good as it could be. The opponent of this change fails to justify her status quo preference by appeal to transition costs, uncertainty, or cognitive limitations. She prefers the status quo, Bostrom and

29. Thanks to Larry Temkin for suggesting these examples.

30. Examples involving love are closer to what Cohen calls personal valuing. But the point of these examples is to draw an analogy between love and value. Even when we have no personal connection to the valuable things which we have reason to preserve, the correct attitude to have is relevantly like the correct attitude in these examples.

Ord reason, because it is the status quo. Her disposition to prefer the status quo (which is perhaps a function of other cognitive biases) is doing the work here, not some object-given reason in favor of the status quo. Conservatism, however, comes to the (partial) rescue. She has, indeed, some reason to prefer the status quo: both departures from the status quo would degrade or destroy some existing thing of value—for example, our existing ways of life, our culture’s self-conception in all its imperfections, or something to that effect. This fact, according to Conservatism, is a reason to prefer the status quo. This reason, of course, may not outweigh the reasons to increase human intelligence—hence the merely partial rescue. Another way to see that Conservatism supports status quo bias is that, on a conservative theory, seemingly irrelevant factors such as existence and loss aversion have normative weight. It is worse, according to Conservatism, to lose an existing thing of value than to fail to gain a thing of equivalent value. And this fact will make the status quo more preferable than we might have initially thought. Conservatism allows many cases of status quo bias to pass Bostrom and Ord’s Reversal Test.

I, therefore, share Jonathan Pugh, Guy Kahane, and Julian Savulescu’s suspicion that Conservatism manifests status quo bias, despite Cohen’s attempt to resist this conclusion. Cohen points to a paragraph in Bostrom and Ord’s article, which he takes to exclude its application to Conservatism.³¹ The paragraph in question restricts Bostrom and Ord’s focus to evaluative beliefs about consequences. Cohen claims that his view is opposed to consequentialist views, according to which our moral obligations are solely a function of the value of consequences, because we have reason not to treat a valuable thing merely as a means to the promotion of value.³² We might, therefore, infer that his conservative principle cannot have implications about the value of consequences. But that inference would be a mistake. Cohen’s nonconsequentialism is compatible with the view that the world is better, in some respect, when we follow Conservatism—just as Rawls’s nonconsequentialism is compatible with the view that the world is better, in some respect, when we follow the principles of justice as fairness. We should follow Cohen in denying that our reasons to preserve valuable things just are reasons to promote value—that is, that our reasons to preserve a valuable object are reducible to the object’s contribution to the value of the state of affairs

31. Jonathan Pugh, Guy Kahane, and Julian Savulescu, “Cohen’s Conservatism and Human Enhancement,” *Journal of Ethics* 17 (2013): 331–54; also see n. 33.

32. Ralf M. Bader, in his review of Cohen’s *Finding Oneself in the Other* (*Notre Dame Philosophical Reviews*, April 2013; <http://ndpr.nd.edu/news/38715-finding-oneself-in-the-other/>), argues that this claim is misleading and that Cohen’s Conservatism is best understood as an axiological view. Here, however, I simply take Cohen’s claim at face value.

in which it exists. But it can still make such a contribution. And the contribution in this case is not simply an additive-aggregationist function, as if a greater number of valuable things would entail a proportionate increase in the value of the outcome. Rather, the added value to the state of affairs comes from our doing something that we have reason to do—namely, preserving existing bearers of value as such.³³

There is independent reason to think that nonconsequentialist principles have implications for the value of consequences. As Larry Temkin argues, acting rightly and doing what one has reason to do may be good-making features of outcomes.³⁴ The extent to which such considerations affect the value of outcomes may, in some cases, be great enough to change the all-things-considered ordinal ranking of options. Consider the following case:

Lifebuoy. Two strangers are drowning. I can prevent both their deaths by taking a lifebuoy from my mother, who would then drown.³⁵

Many people's intuitions about this case would respond to two deontological aspects—namely, my special obligation to my mother, and the distinction between acts and omissions (or the more specific distinction between killing and letting die). But if we focus only on the value of the outcomes and ignore the deontological features, it seems that the world in which I save the two strangers would be at least as good as (and probably better than) the world in which I let them die. We can see this by considering a case with the same features, except that both deaths result from acts and my mother is not involved (i.e., the lifebuoy would come from a third stranger). Unless we reject aggregation and deny that more deaths are worse, we should conclude that the outcome is better if I save the two strangers.

Crucially, however, we may focus on the value of the outcomes while also considering the deontological aspects as good-making features. In *Lifebuoy*, it is plausible that the world in which I let two strangers die is, all things considered, better than the world in which I kill my mother, in virtue of the fact that I act rightly in the former world and wrongly in the latter. One might, of course, admit that rightness is a good-making feature of outcomes and yet doubt my claim that it tips the balance of axiological considerations in *Lifebuoy*. But we can revise the case to

33. Bostrom and Ord ("The Reversal Test") mention this possibility in a footnote (n. 1).

34. Larry Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*, Oxford Ethics Series (Oxford: Oxford University Press, 2012), 205.

35. This example is similar to Temkin's (*ibid.*), but Temkin offers no examples where rightness is a good-making feature that tips the balance of reasons.

make the nondeontological axiological considerations a tie, or close to a tie, such that rightness is an axiological tiebreaker. Consequentialists will maintain that such cases are impossible, because all deontic facts (on their view) hold in virtue of the axiological facts. Since the in-virtue-of relation is asymmetric, the axiological facts cannot also hold in virtue of deontic facts (as they do if rightness is, indeed, a good-making feature of outcomes). But consequentialists of this stripe are unlikely to accept Conservatism in the first place, so they will have already gotten off the boat.

If the rightness of bringing about some outcome can be a good-making feature of outcomes, then rightly preserving things of value may make the status quo better. Cases like *Lifebuoy* suggest that such deontological considerations can, in themselves, have axiological implications. So Conservatism may support status quo bias even if it is a primarily deontological principle.

I have not given a new argument for Conservatism or defended it from objections. My aim has been to explain how, on this independently plausible objective theory, status quo bias is fully rational in some cases. It is fully rational when and because our preference is responsive to our reason to preserve existing things of value. When we do what, according to Conservatism, we have reason to do, this makes the outcome in some way better.

IV. REFERENCE-DEPENDENT PREFERENCES

Suppose we accept Cohen-style Conservatism and agree that status quo bias is sometimes fully rational. There may seem to be a problem with our preferences, which both subjectivists and objectivists would agree is irrational. Suppose that x and y are some valuable things to which Conservatism applies. Suppose further that when x exists in the status quo and y does not, x is strictly preferred to y : $x \succ y$. When y exists in the status quo, we have the opposite preference: $y \succ x$. But we cannot rationally prefer x to y and prefer y to x . How should we understand our preferences in a way that does not have this result?

In behavioral economics, status quo bias is most commonly understood in terms of preferences that change over time. According to Tversky and Kahneman, our preferences depend on our reference point or endowment.³⁶ When our reference point is x , our choice behavior is explained by the preference relation \succ_x . When our reference point is y , our choice behavior is explained by the preference relation \succ_y . So, instead of one preference relation \succ on which $x \succ y$ and $y \succ x$, Tversky and

36. Amos Tversky and Daniel Kahneman, "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics* 106 (1999): 1039–61.

Kahneman would say that $x \succ_x y$ and $y \succ_y x$, which need not be irrational. On this view, we have a family of preference relations, and we change between them based on the reference point. The changing preferences model allows us to avoid the result that we have inconsistent preferences at any given time.³⁷

That is a descriptive explanation of the phenomenon of status quo bias. My aim in this section is to discuss this model's normative implications. I first discuss an objection to changing preferences, which may give rise to an objection to Conservatism and status quo bias. The objection is that permitting changing preferences seems to support "sour grapes" preferences, where we come to prefer something because we can no longer get what we initially wanted. I explain how we can solve the problem without ruling out status quo bias. I then try to offer some independent justification for the idea that our preferences may rationally depend on different reference points for different comparisons. I argue that this idea may be justified by a powerful view about outcome goodness, which may help to vindicate the conservative disposition.

A. *Adaptation and Sour Grapes*

We often change our preferences to what we are likely, rather than unlikely, to get. These adaptive preferences often fail to align with what we have most reason to want. Amartya Sen, for example, notes that subjugated agents (e.g., "the hopeless beggar" and "the dominated house-

37. There are two main alternatives to Tversky and Kahneman's view. One is to give a more fine-grained individuation of the things preferred. Frederic Schick, in "Status Quo Basing and the Logic of Value," *Economics and Philosophy* 15 (1999): 23–42, claims that we value propositions, which depend on how we see or understand outcomes. So, whereas Tversky and Kahneman see us as preferring candy to a mug, where the preference is indexed to some reference point ($c \succ_c m$), this alternative might see us as preferring candy-when-we-have-candy to a mug-when-we-have-candy ($c_c \succ m_c$). But, in its normative implications, Schick's view permits not just status quo bias, but almost all framing effects. We need an objective theory of reasons to determine when we should see outcomes as distinct propositions in a way that justifies a preference between them. This echoes John Broome's claim, in *Weighing Goods* (Oxford: Blackwell, 1991), 105, that substantive normative assumptions are needed to individuate alternatives so that coherence requirements can constrain preferences. My argument in Sec. II, therefore, provides a necessary supplement to Schick's view. Another alternative to Tversky and Kahneman's view, due to Michael Mandler ("Status Quo Maintenance Reconsidered: Changing or Incomplete Preferences?" *Economic Journal* 114 [2004]: F518–F535), understands status quo bias in terms of incomplete weak preferences. An agent's weak preference relation is incomplete if there are some outcomes between which she has no preference and is not indifferent. Mandler sees status quo bias as an example of how an agent with incomplete preferences can, in a principled way, avoid missing foreseen opportunities to obtain strictly preferred final outcomes. This kind of view appeals to the good effects of certain status quo-maintaining policies (e.g., to deviate from the status quo only if the alternative is at least weakly preferred), which is very different from the kind of justification I have sketched.

wife”) may evolve distorted preferences to cope with their situations.³⁸ In the extreme case, subjugated agents may even come to prefer their subjugation. Dale Dorsey claims—plausibly, I think—that these agents’ preferences for their subjugation lack authority over their well-being.³⁹ Some preference has authority over an agent’s well-being just in case having this preference is a sufficient condition for the preferred option being all-things-considered better for the agent than its alternative. Dorsey also claims that these adaptive preferences lack authority because they manifest status quo bias.⁴⁰ If that is true, then we may have strong reason to reject Conservatism, given the connection between Conservatism and status quo bias. The question, then, is how to rule out sour grapes preferences without ruling out status quo bias and reference-dependent preferences more generally.

Since I believe that status quo bias is not inherently irrational, I need an alternative explanation for Dorsey’s claim that the adaptive preferences in Sen’s examples lack authority over their agents’ well-being. Fortunately, we need not look far for such an explanation. We should accept an objective theory of the rationality of desires and preferences, and we can deny that irrational preferences have authority over their agents’ well-being.⁴¹ I have already argued that subjective theories of rational preferences cannot plausibly condemn status quo bias as irrational, so an objective theory ought to be common ground here. On an objective theory, we can plausibly insist that it would be irrational for the dominated housewife to prefer her domination: she has reasons to want not to be dominated, which are given by the properties of domination. Combined with the plausible principle that irrational preferences lack authority over their agents’ well-being, this view explains why the adaptive preferences noted by Sen lack authority over their agents’ well-being.

Dorsey, however, claims that the irrationality of preferring what is objectively worse for us does not rule out all objectionable preferences

38. Amartya Sen, *On Ethics and Economics* (Oxford: Blackwell, 1987), 45–46.

39. Dale Dorsey, “Preferences, Welfare, and the Status-Quo Bias,” *Australasian Journal of Philosophy* 88 (2010): 535–54.

40. It is not clear to me whether Dorsey finds status quo bias irrational. He often calls it “problematic,” but this may only mean that it is nonauthoritative. Dorsey sometimes says that the process of, or attempt at, preference adaptation is rational, but I don’t know whether he thinks the preferences that result from this process or attempt are thereby rational.

41. The objective/subjective distinction for theories of well-being may cut across the objective/subjective distinction for theories of reasons and rationality. For example, an objective theory of reasons, on which we have object-given reasons to care about our future well-being, could be combined with a desire-fulfillment theory of well-being. See Chris Heathwood, “Desire-Based Theories of Reasons, Pleasure, and Welfare,” *Oxford Studies in Metaethics* 6 (2011): 79–106.

through adaptation. Dorsey considers an example in which the agent's preference seems to be a function of sour grapes:⁴²

The Painter. Erin, at t_{-1} , preferred a life as a great dancer to the life of a great painter and dedicated years to becoming a dancer. But, after much rejection and failure, she decided to become a painter. To avoid continued frustration and regret, she alters her preferences away from dancing and toward painting, which is her new status quo. Suppose she is successful: Erin, at time t , prefers the life of a great painter to the life of a great dancer.

Dorsey claims that despite Erin's preference at t for the life of painting, being a great painter is not all-things-considered better for her at t than being a great dancer. On this view, Erin's new preference is not authoritative over her well-being. Why not? Her new preference's lack of authority is not explained by the irrationality of preferring what is objectively worse for her. Being a great painter does not seem objectively worse for Erin than being a great dancer. Dorsey thinks this lack of authority is instead explained by the fact that her preference manifests status quo bias.⁴³

Dorsey argues that preferences which manifest status quo bias are not authoritative because they violate the following principle:

Independence: Whether x would be better for a than y at t is independent of whether x rather than y is more likely to occur.⁴⁴

This principle is incompatible with the authority of preferences which exhibit status quo bias, because such preferences depend on whether the preferred state of affairs occurs (or is likely to occur). But, importantly, Independence is also incompatible with uncontroversially good reasons to prefer the status quo. Consider transition costs, for example. The fact that using the metric system would be worse for the United States than maintaining its customary measurement system does depend on the fact that the United States currently uses the latter. The

42. Dorsey, "Preferences," 538.

43. In order to understand this case as an instance of status quo bias, as Dorsey does, we need to understand the relevant status quo at least partly in terms of the agent's expectations about the future. Erin's preference for painting is motivated by the fact that she is more likely to satisfy it than to satisfy a preference for dancing. Status quo bias, so understood, is a bias in favor of actual states of affairs, not merely a bias in favor of presently obtaining states of affairs. Cohen's Conservatism should also be understood in this way (see Cohen, "Rescuing Conservatism," 165–66).

44. Dorsey, "Preferences," 540. "More likely," I assume, has to do with the probabilities relative to a 's credal state, not with the objective chances.

upshot is that Independence is too strong, because it rules out uncontroversially good reasons to prefer the status quo.⁴⁵

A more controversial, but no less important, problem with Independence has to do with alienation.⁴⁶ Which is better for me, all things considered: to be a great basketball player or to be a decent philosopher? Being a great basketball player (in any world that is sufficiently similar to the actual world in other relevant respects) would require me to have attitudes that differ greatly from my actual attitudes—for example, more athletic ambition and less love of philosophy than I have. From what point of view should we assess which of these lives, with the attitudes required to lead them, would be better for me? I am inclined to think that my status quo attitudes can legitimately influence which life is better for me. According to Independence, however, the correct evaluation must transcend my present or expected point of view, so I should abstract from my love of philosophy and lack of athletic ambition. But why should the evaluation of what is better for me be so completely disconnected from my actual attitudes? Which of two states a person has most reason to prefer for her own sake often seems to depend on the temporal or modal point of view from which she evaluates these states—that is, in Tversky and Kahneman’s ideology, the person’s reference point. For example, disability may lead a person to have certain valuable projects, relationships, and attitudes. From her present perspective, which is shaped by these projects, relationships, and attitudes, she may reasonably regard her disability as better (for her) than how she would have regarded it from a very different point of view.⁴⁷ Such a different point of view, however, may be so alien to her that it seems irrelevant to what she has most reason to prefer for her own sake. I do not find Independence so independently plausible as to rule out such reference-dependent preferences as lacking authority.

45. Perhaps Independence can be restricted so that it wouldn’t be open to this objection. I am not confident, however, that a narrower interpretation would achieve Dorsey’s aim of condemning status quo bias.

46. Thanks to an associate editor of *Ethics* for suggesting that I discuss this problem here. A similar worry is also expressed by Rawls, *A Theory of Justice*, 152, in his contrast between “bare-persons” and “determinate-persons.”

47. See Robert Adams’s discussion of Helen Keller in “Existence, Self-Interest, and the Problem of Evil,” *Noûs* 13 (1979): 53–65. I do not share Adams’s judgment that it is unreasonable for Helen Keller to wish that she had never been deaf or blind. But I am inclined to accept Adams’s claim that we may reasonably be glad that we have our actual partners, children, and lives, rather than having had others that (when evaluated independently of our actual points of view) would have made things go even better. See also Elizabeth Barnes, “Disability and Adaptive Preference,” *Philosophical Perspectives* 23 (2009): 1–22; and Elizabeth Harman, “‘I’ll Be Glad I Did It’ Reasoning and the Significance of Future Desires,” *Philosophical Perspectives* 23 (2009): 177–99.

Return to The Painter. Is Erin's preference authoritative? One might think that Erin's preference is authoritative on the grounds that her change of preference is good for her. Martha Nussbaum comes close to endorsing this kind of view when she claims that we should often discourage people's unrealistic aspirations in order to prevent disappointment and that it therefore seems rational to adjust one's preferences accordingly.⁴⁸ Her own example is a short child who wishes to become the world's best basketball player, or a bad singer who hopes to become the best opera singer.

Nussbaum may be right that adaptation is often good for us. But this does not entail that the resulting preferences are rational or authoritative: just because it would be good for us to have some preference doesn't imply that the preferred object is good. Even if it would be better for me to prefer the life of a philosopher to the life of the world's best basketball player, the life of a philosopher might not be better for me than the life of the world's best basketball player.

This distinction is compatible with an objective theory of rational preferences. I assume, following Parfit, that the rationality of our attitudes is a function of object-given reasons, which derive from the contents or objects of those attitudes, rather than state-given reasons, which derive from the properties of the attitudes themselves.⁴⁹ For example, the rationality of believing p is a function of our evidence for p , not of the benefits of being in the state of believing p . These benefits may make it rational to want to believe p , but they don't make it rational to believe p . Similarly, Nussbaum's advantages of adaptation fundamentally derive not from the contents or objects of adaptive preferences, but from the properties of the attitudes or the state of preferring. This fact may make it rational to want to prefer what is likely, rather than unlikely, to occur. But it doesn't make it rational to prefer what is likely, rather than unlikely, to occur, and it does not make likely outcomes better for us than unlikely outcomes.⁵⁰

My view is that Erin's preference in The Painter may or may not be rational and authoritative once she becomes a painter. In accordance with Conservatism, Erin has some reason to prefer to maintain the status quo. Suppose that Erin learns that she could, in fact, succeed at dancing. If, because of this information, she ends her career as a painter

48. Martha Nussbaum, *Women and Human Development: The Capabilities Approach* (Cambridge: Cambridge University Press, 2000), 137–38.

49. Parfit, *On What Matters*, vol. 1, sec. 5. Those who reject Parfit's distinction, such as Mark Schroeder ("The Ubiquity of State-Given Reasons," *Ethics* 122 [2012]: 457–88) can instead understand my point here in terms of their favored account of the "right" and "wrong" kinds of reasons.

50. Dorsey, "Preferences," 545 n. 14, cites an early statement of Parfit's distinction, but he applies it only to ideal adviser theories of value.

and decides to return to her initial goal of becoming a great dancer, she might lose some existing thing of value—namely, her self-conception as a painter. But this reason may be very weak, so it may not outweigh her reasons to prefer a career as a dancer.

This explanation gives a plausible verdict in other cases of preference adaptation. Cohen claims that we ought to accept some things as given, because “the attitude of universal mastery over everything is repugnant, and, at the limit, insane.”⁵¹ The things taken as given seem to include valuable objects and states of affairs. David Wiggins, for example, suggests that unless we accept certain natural limits on our desires, our sense of omnipotence will make the world less meaningful.⁵² And Michael Sandel argues that the “Promethean assault on the given” will undermine our appreciation for life.⁵³ I shall not assess these claims here, but I mention them because they might seem to support the rationality of preference adaptation to our given circumstances. But even if Cohen, Wiggins, and Sandel are right, Dorsey can plausibly maintain that we are too eager to relinquish mastery and to concede to the given in many cases of adaptation (as he thinks Erin is in *The Painter*). Perhaps we have some reason to accept what is given—for example, our families, our given natures, or our values—but this disposition can become excessive and irrational.

B. The Status Quo as a Reference Point

I now turn to the more general question of whether it can be rational to change our preferences depending on our reference point. Cohen-style conservatives tend to use the status quo as a reference point for comparison. Instead of comparing each option with respect to each available alternative, conservatives often care more about comparisons to the status quo. Anticonservatives view this tendency as irrational and would reject the assumption that one state of affairs has a privileged status when it comes to the comparison of alternatives. That view may explain why, in the experiments that test for status quo bias, it seems irrational for the agent’s choices in neutral framings and in status quo framings to differ so significantly. In this section, I sketch a potential defense of our conservative disposition to compare options with respect to the status quo. This defense may help to vindicate the rationality of reference-dependent preferences, which Tversky and Kahneman use to explain status quo bias.

51. Cohen, “Rescuing Conservatism,” 149.

52. David Wiggins, *Sameness and Substance Renewed* (Cambridge: Cambridge University Press, 2001), 242.

53. Michael Sandel, *The Case against Perfection* (Cambridge, MA: Harvard University Press, 2007).

Larry Temkin offers some powerful reasons to think that *all-things-considered better than* might not be a transitive relation. If this relation is not transitive, then it is possible that some outcome O_1 is better (all things considered) than some alternative O_2 , which is in turn better than O_3 , and yet O_1 is not better than O_3 . Many of Temkin's arguments are spectrum arguments, which present a series of outcomes, lives, or experiences, where each alternative varies with respect to more than one dimension—for example, duration, intensity, or frequency. One such argument begins with a very long life containing two years of excruciating torture.⁵⁴ This life seems better than an equally long life with four years of slightly less intense torture, and this second life seems better than a third, equally long life with eight years of torture slightly less intense than in the second life. Iterations of this reasoning (slightly lower intensity with double duration) yield a life with a very long duration of mild discomfort, compared to an equally long life with a much longer duration of discomfort almost as bad—in this case, one additional mosquito bite per month. But, while the former may be better than the latter, it seems that the extra mosquito bite per month is better than the two years of horrible torture. If *better than* is a transitive relation, then a series of steps for the worse would make the longest ordeal also the worst.

This kind of case leads Temkin to defend an Essentially Comparative View of Outcome Goodness, according to which the goodness of an outcome depends on the alternatives with which it is compared. If the Essentially Comparative View is correct, then this fact has radical implications for practical reasoning. Temkin now agrees, for example, that the Essentially Comparative View is incompatible with the claim that a world filled with undeserved suffering is bad, or worth wanting to avoid for its own sake.⁵⁵

Unless we are very confident that Temkin's arguments are unsound, we should consider some ways to maintain the action-guiding importance of normative evaluations on an Essentially Comparative View. One option is to rely on pairwise comparisons to some relevant baseline or reference point. When it comes to practical reasoning, the status quo

54. This argument appears in sec. 5.3 of Temkin's *Rethinking the Good* and is based on Stuart Rachels's spectrum argument from torture to headaches in "A Theory of Beneficence" (undergraduate thesis in philosophy, politics, and economics, University of Oxford, 1993). See also Larry Temkin, "A Continuum Argument for Intransitivity," *Philosophy and Public Affairs* 25 (1996): 175–210.

55. This implication may seem to undermine the conservative reason to preserve things that are intrinsically valuable. But the Essentially Comparative View only rules out the intrinsic value of outcomes. Undeserved suffering can be bad even if a world filled with undeserved suffering cannot be. That seems to be the view of nonconsequentialists who deny that states of affairs have intrinsic value.

might be a relevant baseline: should we maintain the way things are now, or should we opt for some kind of change? If Temkin is right, then there may be no coherent ranking of alternatives with respect to each other. But each alternative may be better or worse than the status quo to a certain extent. And these pairwise comparisons to the status quo may determine what we ought to do. If an outcome's goodness depends on the alternatives with which it is compared, and if every outcome might be worse than some alternative, then moral evaluation can guide action only if we privilege some option as a reference point for comparison. And what better option than the status quo? My suggestion is a reference-dependent value relation, where the status quo is the reference point to which the action-guiding value relation is indexed.

In some cases, we seem to make important comparisons among options that are not the status quo. Temkin, for example, supposes that we are on a search committee that aims to fill a vacancy in employment, but each applicant is better than the status quo, in which we have no one.⁵⁶ The applicants, however, might each be better than the status quo to differing extents. We should choose the applicant that is better than the status quo to the greatest extent, even though this applicant may (on Temkin's view) be worse than some alternative applicant. We may have reason to regret choosing this applicant, since there is a better alternative relative to our choice. But that fact, Temkin thinks, may be true of every applicant.⁵⁷ There may be some option which, given our starting point, we have most reason to choose, even though every option would leave us with some reason for regret.

This kind of view may help to explain why states of affairs seem to have intrinsic value on an Essentially Comparative View. When we say that an outcome is good, or worth wanting for its own sake, we mean that it is better than, or preferable to, some reference point to which we compare outcomes of that kind. Consider whether or not it would be good if lots of people existed with lives that were barely worth living. We may be comparing this outcome with respect to a state of affairs in which no one ever exists: it would seem good if lots of people existed with lives

56. Temkin, *Rethinking the Good*, sec. 13.3.

57. One of Temkin's most plausible examples of this possibility is an affirmative action case, because affirmative action considerations may be relevant and significant for some pairwise comparisons, but not for others (*Rethinking the Good*, 212). Suppose that affirmative action considerations count in favor of choosing Applicant A over Applicant B, but that they are irrelevant in the comparisons between Applicant B and Applicant C and between Applicant A and Applicant C. That is likely, for example, if affirmative action is justified by historical relations between particular races or ethnicities. If the applicants score equally well in all ways unrelated to affirmative action, then our pairwise-comparative judgments in this case may violate the transitivity of the *equally as good as* relation, which would leave little room for maintaining the transitivity of *better than*.

that were barely worth living just in case that would be better than a state of affairs in which no one ever exists. Perhaps we compare the status quo with respect to some historical expectation: the way things are now seems good to us just in case it is better than the way things have usually been. More generally, some fact is a reason to want some state of affairs for its own sake (on the view we are considering) just in case this fact is a reason to prefer this state of affairs to some relevant baseline. Without such a reference point, we might find it hard to explain our basis for thinking that some state of affairs is good: “Good compared to what?” one might ask. And the status quo baseline makes a plausible connection between the value of possible states of affairs and reasons for action: when we deliberate, we consider departures from our starting point, and we ought to choose the best option compared to that starting point.

Of course, even if there is good reason to make decisions based on how alternatives compare relative to the status quo, there may also be good reason to want to know how alternatives compare to other alternatives besides the status quo. So even if Temkin’s arguments lend some support to this proposal, it does not follow that the status quo baseline view solves the problems Temkin discusses. I am not certain that we should maintain transitivity in light of Temkin’s arguments, so it seems to me that a status quo baseline view is worth considering. If Temkin’s arguments provided some grounds for the conservative disposition to compare alternatives with respect to the status quo, that would strike me as an interesting result.

V. CONCLUSION

I have argued that status quo bias is, in some cases, fully rational on a plausible objective theory of preference rationality. On this conservative theory, we have reason to preserve valuable things, even when we can replace them with things of equal or greater value. This justification depends in no way on the good effects of status quo bias. But I have also defended status quo bias on subjectivist grounds, by rejecting some ways in which status quo bias might violate constraints on our preferences. And I have argued that the conservative disposition to compare options with respect to our status quo reference point may be rational.

Nonetheless, we should be careful of status quo bias. Although status quo bias may be rational when we give the conservative principle its due weight, this bias may lead us to give this principle too much weight—that is, to prefer the status quo too often and to use this principle as rationalization. We should be especially careful of this bias when the status quo perpetuates some kind of injustice or harm. Injustice and harm have disvalue, so Conservatism does not count in favor of preserv-

ing them. Many states of affairs, however, involve some things that are good and others that are bad—for example, in Cohen's example, slavery and the pyramids in ancient Egypt. I believe that, in most cases, injustice and harm matter more than the preservation of existing things of value. Doing what we ought to do may, therefore, often leave us with reasons to regret the loss of valuable things.