

## Ethics and the Limits of Armchair Sociology\*

Brendan de Kenessey  
University of Toronto

*The Journal of Philosophy*, 2024  
Pre-publication draft; please cite [published version](#)

**Abstract:** Contractualism and rule consequentialism both hold that whether a moral principle is true depends on what would happen if it were generally adopted as a basis for conduct. This paper argues that any theory with this feature faces a profound epistemic problem. The question of what would happen if different moral principles were generally adopted is a complex empirical question, comparable in difficulty to the question of what would happen if a nation adopted different laws, or if humanity had evolved different traits. Reflection on the epistemic demands of this question shows that we have no clue what would happen if different moral principles were generally adopted. This leads to a surprising result: we have no clue what contractualism and rule consequentialism imply about what actions are morally right and wrong. The only way for these theories to avoid cluelessness is to test principles on groups small enough to be epistemically tractable, which requires accepting an implausibly extreme form of moral relativism. I conclude that we must reject contractualism, rule consequentialism, and any other moral theory that entails that the truth of a moral principle depends on what would happen if it were generally adopted.

Several leading theories in normative ethics pursue a common strategy: to determine whether a moral principle is true, they ask what would happen if it were generally adopted. It is wrong to lie, according to these theories, because if everyone lied whenever they wanted then no one would

---

\* For generous and helpful comments on drafts of this paper, I thank David Barnett, Thomas Byrne, Sanford Diehl, Andrew Franklin-Hall, Caspar Hare, Daniel Muñoz, Luis Oliveira, Bernhard Salow, Ben Schwan, Kieran Setiya, Jonas Vandieken, an audience at the Massachusetts Institute of Technology, and (especially) a reviewer for this JOURNAL.

This paper's title echoes that of Bernard Williams's classic *Ethics and the Limits of Philosophy* (Cambridge: Harvard University Press, 1986). The similarities mostly end there, though Williams was also opposed to generalization-style moral theories for his own reasons.

trust others' testimony. It is wrong to violate bodily rights, even for the greater good, because if everyone felt licensed to do so then no one would feel secure in their bodily autonomy.

The most famous theory of this kind is Kant's Formula of Universal Law,<sup>1</sup> but I will focus on two more contemporary theories: *contractualism* and *rule consequentialism*. Both theories hold that whether a moral principle is true depends on what would happen if it were generally adopted as a basis for conduct. Rule consequentialism says that the true moral principles are those that would have the best consequences if generally adopted.<sup>2</sup> Contractualism, in T. M. Scanlon's canonical formulation, says that the true moral principles are those that no one could reasonably reject as a basis for the general regulation of behavior.<sup>3</sup> In turn, whether a principle can be reasonably rejected depends in large part on what would happen if it were generally adopted.<sup>4</sup>

In other words, contractualism and rule consequentialism both explain moral principles by appeal to counterfactual claims about what would happen if they were generally adopted. I shall call these *generalization counterfactuals*: counterfactuals of the form "if such-and-such moral principles were generally adopted, then the world would be thus-and-so." And I shall use the term *generalization theory* for any moral theory that implies that which moral principles are true

---

<sup>1</sup> Immanuel Kant, *Groundwork of the Metaphysics of Morals*, translated by Mary Gregor and Jens Timmerman (Cambridge: Cambridge University Press, 1785/2012).

<sup>2</sup> See, for instance, Richard Brandt, *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1979), and Brad Hooker, *Ideal Code, Real World* (Oxford: Oxford University Press, 2002).

<sup>3</sup> T. M. Scanlon, "Contractualism and Utilitarianism," in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982), pp. 103–28, at p. 110; and T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998).

<sup>4</sup> See Scanlon, *What We Owe*, *op. cit.*, p. 204.

depends on what would happen if different moral principles were generally adopted. On their usual interpretations, contractualism and rule consequentialism are both generalization theories.

This paper argues that any generalization theory that also accepts *moral universalism*—the claim that the same basic moral principles apply to all persons—faces a profound epistemic problem. To get a sense of the problem, consider a counterfactual in a different domain: what would have happened if humanity had never evolved religious belief? Would the lack of belief in an overseeing deity have stunted the development of legal and moral systems? Or would humans have been more peaceful and rational without the distorting fervor of religion? How would the absence of religion have altered the development of politics, science, and technology, norms of gender and sexuality, or geographical patterns of trade, migration, and conquest?

Reflecting on this counterfactual, two things become clear. First, the question of what would have happened if religion had not evolved is an *empirical* question. Its answer depends on contingent facts about human psychology and sociology that cannot be determined *a priori*. Second, we do not have nearly enough data to know, even approximately, what would happen in this counterfactual scenario. The scope of the counterfactual intervention is so huge, and the sociological mechanisms it involves are so complex, that all we can do is speculate. There are many possible answers to the question “What would happen if religion had never evolved?” that are all equally compatible with our evidence.

Compare a generalization counterfactual: what would happen if most of humanity accepted a moral principle requiring one to never lie under any circumstances? Without the need to protect against lying, would we have developed robust judicial systems, or the practice of investigative journalism? Would the “never lie” principle make dictatorships more likely by giving aspiring autocrats a trusting populace, or less likely by making it harder to control information and spread

propaganda? How would the absence of lying have altered the development of politics, science, and technology, norms of gender and sexuality, or geographical patterns of trade, migration, and conquest?

These questions are just as opaque as the parallel questions about religion. Like our religious counterfactual, a generalization counterfactual posits a sweeping change in human attitudes and dispositions across history. The question of what would result from this change is an empirical matter that depends on contingent features of human psychology and sociology. And, as in the religion case, we do not have nearly enough data to answer it. The scope of generalization counterfactuals is so huge, and the mechanisms they involve are so complex, that all we can do is speculate. There are many possible answers to the question “What would happen if the ‘never lie’ principle were generally adopted?” that are all equally compatible with our evidence. In sections II and III, I shall argue that we have no clue which universalist generalization counterfactuals are true. This leads me to the surprising conclusion that we have no clue what universalist contractualism and rule consequentialism imply about what actions are right and wrong.<sup>5</sup>

---

<sup>5</sup> A few authors express similar epistemic worries in passing: see James Griffin, *Value Judgment* (Oxford: Clarendon Press, 1996), pp. 106–7; Thomas Pogge, “What We Can Reasonably Reject,” *Philosophical Issues* XI (2001): 118–47, at p. 135; Rob Lawlor, “Hooker’s Ideal Code and the Sacrifice Problem,” *Social Theory and Practice* XXX, 4 (2004): 583–87, at pp. 585–86; Alison McIntyre, “The Perils of Holism: Brad Hooker’s *Ideal Code, Real World*,” *Philosophical Issues* xv (2005): 252–263, at pp. 255–60; and Holly Smith, “Measuring the Consequences of Rules,” *Utilitas* XXII, 4 (2010), 413–33, at pp. 427–28. So the epistemic problem that this paper raises has not gone completely unnoticed. I believe, however, that the problem’s gravity and scope have yet to be widely appreciated.

This paper’s argument is also similar to, and borrows the term “clueless” from, James Lenman’s argument that we are clueless what *act* consequentialism implies about what we ought to do (James Lenman, “Consequentialism and Cluelessness,” *Philosophy & Public Affairs* XXIX, 4 (2000): 342–70). But my argument is independent from

However, there is a way out. Universalist generalization counterfactuals are so epistemically impenetrable because they are so broad in scope: they ask what would happen if a principle were generally adopted across all of humanity throughout all time. A natural solution is to test principles on smaller groups. The result is a *relativist* generalization theory: a theory that takes different basic moral principles to apply to different groups, depending on what would happen if those principles were generally adopted within each group.

---

Lenman's: one could be persuaded that we are clueless about universalist generalization counterfactuals without being convinced that we are clueless about the consequences of individual actions, or vice versa. In particular, our cluelessness about generalization counterfactuals does not depend on the "butterfly effects" central to Lenman's argument, such as a prehistoric action bringing about the birth of Hitler.

There is one more objection in the recent literature that bears some resemblance to this paper's argument. The objector invites us to suppose that a principle's general adoption would trigger some otherwise inert process that produces either wonderful or disastrous effects: say, a gremlin who unleashes devastation if everyone accepts a promise-keeping principle (Gideon Rosen, "Might Kantian Contractualism be the Supreme Principle of Morality?," *Ratio* XXII, 1 (2009): 78–97; for similar arguments, see Jussi Suikkanen, "A Dilemma for Rule-Consequentialism," *Philosophia* XXXVI, 1 (2008): 141–50; and Abelard Podgorski, "Wouldn't it be Nice? Moral Rules and Distant Worlds," *Noûs* LII, 2 (2018): 279–94). Generalization theories appear to predict that if this gremlin did exist, then it would be wrong to keep your promises even when doing so would not trigger the gremlin's wrath—which seems absurd. However, the gremlin objection does not question the generalization theorist's claims about the direct effects of a principle's general adoption on human society; it simply adds an extra effect that skews the calculus. This offers the generalization theorist a clear reply strategy: find some way of excluding the effects of gremlins (and similar interventions) when evaluating principles (see Rosen, "Might Kantian," *op. cit.*, pp. 88–90). The argument offered here cuts deeper, as it calls into question the generalization theorist's claims about the *normal, direct* effects of principles' general adoption.

I argue in section V that this option comes with steep theoretical costs. For those like me who find moral relativism flatly implausible, accepting relativism is itself a cost. But setting this aside, my main argument will be that, for the relativist strategy to succeed in making generalization counterfactuals epistemically tractable, the groups on which moral principles are tested must be quite small—in the ballpark of the population of Switzerland for a ten-year span. Thus, to avoid the epistemic problem, generalization theorists must adopt a radical form of relativism, according to which moral principles vary across small distances in space and time.

Thus contractualists, rule consequentialists, and any other generalization theorists face a dilemma. If they accept moral universalism, they encounter the result that we have no clue what principles their theories endorse, as we have no clue what would happen if all of humanity adopted different moral principles. If they try to avoid cluelessness by testing principles on smaller groups, they are committed to an implausibly fine-grained moral relativism and all the theoretical problems that come with it. As I find neither horn of this dilemma acceptable, I conclude that we should reject all generalization theories, and thus reject contractualism and rule consequentialism.

I conclude (section VI) by observing that contractualists have a third option: they can abandon the idea that whether a principle can be reasonably rejected depends on what would happen if it were generally adopted. This would sidestep the dilemma, but at the cost of depriving contractualism of much of its explanatory power.

Although I focus on contractualism and rule consequentialism, the true target of this paper is the broader paradigm of explaining moral principles by appeal to the effects of their general adoption. Any theory that employs this strategy—that is, any generalization theory—is vulnerable to the argument to come. If whether a maxim can be willed to be a universal law depends on contingent facts about what would happen if that maxim were generally adopted, then Kant's

Formula of Universal Law is subject to the arguments of this paper.<sup>6</sup> If what principles governing the basic structure of society it would be rational to select from behind the veil of ignorance depends on what would happen in a society that adopted those principles, then Rawls's theory of justice is subject to the arguments of this paper.<sup>7</sup> I do not have space to argue that Kant and Rawls should be interpreted as generalization theorists, but this illustrates how the argument might generalize. If my arguments are sound, then we must reject one of the most venerable and popular explanatory strategies in normative ethics—that of asking “what if everyone did that?”<sup>8</sup>

---

<sup>6</sup> Kant, *Groundwork*, *op. cit.*

<sup>7</sup> John Rawls, *A Theory of Justice* (Cambridge: Belknap Press, 1971).

<sup>8</sup> However, asking “what if everyone did that?” is a pervasive part not just of moral philosophy, but of folk moral thought. One might chide someone who jumps the subway turnstile by saying, “If everyone did that, there would be no funding for subways.” Am I really saying that this ordinary style of moral reasoning should be rejected? Not always: so long as this reasoning is small-scale and non-fundamental, it is not subject to my objection. By small-scale, I mean constrained to a local context: we are talking about people paying for the subway of *this* city, for instance, not people contributing to public goods throughout history. More importantly, I think ordinary “What if everyone did that?” reasoning is usually non-fundamental, in that it is meant to highlight a more basic feature that makes an action wrong—often that it unfairly free rides on the efforts of others. It is irrelevant that we do not really know what would happen if nobody paid for the subway (perhaps the government would fund it through taxes instead). The real point is counterfactual-independent: that those who refuse to pay are taking advantage of those who do. Folk employments of “What if everyone did that?” are thus usually on safer epistemic footing. But I maintain that appeal to what would happen if everyone adopted a moral principle cannot serve as a foundational basis for our ethical theory. Thanks to a reviewer for raising this point.

## I. SETTING THE STAGE

*Contractualism* says that we are morally required to act in accordance with the set of moral principles that no person could reasonably reject as a basis for the general regulation of behavior.

*Rule consequentialism* says that we are morally required to act in accordance with the set of moral principles the general adoption of which would produce the best consequences.

Both theories have a *two-level* structure. In the first level, we determine whether an act is morally required by asking whether it is required by certain special principles. In the second level, we determine which moral principles are “special,” that is, which principles apply in the first level, by testing them against some principle selection criteria.<sup>9</sup> Rule consequentialism’s principle selection criterion is simple: the set of moral principles we ought to obey is the set that would produce the best consequences if generally adopted.

Contractualism’s principle selection criterion is more complex. Rather than ranking principles’ consequences impersonally, contractualists consider each person individually, asking what reasons they have to accept or reject a principle. We then determine whether a person can reasonably reject a principle by asking whether any other individual has a stronger objection to its alternatives. A set of principles meets the contractualist’s principle selection criteria, and is thus the set we ought to obey, just in case no individual is in a position to reasonably reject any principles in the set.

As described, contractualism does not yet appeal to generalization counterfactuals. But most contractualists assume that one of the main considerations that determines whether a principle

---

<sup>9</sup> Most generalization theorists apply their test to *sets* of principles, but I will often loosely speak as if we are testing a single principle.



can be reasonably rejected is how that principle's general adoption would affect each individual.<sup>10</sup> Scanlon is explicit: "An assessment of the rejectability of a principle *must take into account the consequences of its acceptance in general*, not merely in a particular case that we may be concerned with."<sup>11</sup>

The arguments of this paper apply equally to contractualism and rule consequentialism, since they are based on these theories' dependence on generalization counterfactuals. I shall thus mostly ignore the distinction between contractualism and rule consequentialism and phrase my arguments as targeting the umbrella category of *generalization theories*.

A crucial question for any generalization theory is: what is involved in a principle's being "generally adopted"? Answering this question precisely requires making several theoretical choices. Some widely discussed choice points do not affect the argument of this paper, such as: what psychological states are involved in "adopting" a principle?<sup>12</sup> What proportion of the population should we imagine adopting the principle?<sup>13</sup> Is a disposition to accept the principles

---

<sup>10</sup> For examples of contractualists invoking generalization counterfactuals, see Scanlon, *What We Owe*, *op. cit.*, p. 200, 203, 205, 223; Rahul Kumar, "Defending the Moral Moderate: Contractualism and Common Sense," *Philosophy & Public Affairs* XXVIII, 4 (1999): 275–309, at pp. 298–299; Nicholas Southwood, *Contractualism and the Foundations of Morality* (Oxford: Oxford University Press, 2010), p. 161; and Derek Parfit, *On What Matters: Volume 1* (Oxford: Oxford University Press, 2011), p. 363, 385.

<sup>11</sup> Scanlon, *What We Owe*, *op. cit.*, p. 204. My emphasis.

<sup>12</sup> See Hooker, *Ideal Code*, *op. cit.*, pp. 75–80.

<sup>13</sup> See *ibid.*, pp. 80–85; Smith, "Measuring," *op. cit.*; and Parfit, *On What Matters*, *op. cit.*, pp. 308–320.

simply hard-wired into our biology, or do the principles need to be taught?<sup>14</sup> For our purposes it does not matter how we answer these questions. As far as I can see, the answers we choose make no significant difference to the epistemic difficulty of generalization counterfactuals.

Let me now turn to two interpretive questions that do affect the argument to come. First, does our generalization theory evaluate principles based on the consequences their general adoption would have *in fact*, or based on the *expected* consequences of the principles' adoption given a body of evidence? Call the former an *objective* generalization theory, and the latter a *subjective* generalization theory. Both views are defended in the literature: Derek Parfit and T. M. Scanlon endorse objectivism, while Richard Brandt and Brad Hooker opt for subjectivism.<sup>15</sup>

The second question is whether our generalization theory holds that the same basic moral principles apply to all persons (*moral universalism*), or that different basic moral principles apply to different groups of persons (*moral relativism*). This question is important for our purposes because it affects the scope of generalization counterfactuals. I assume that the group on which we test a moral principle includes all and only the people to whom that principle applies (I defend this assumption in section IV). Given this assumption, a universalist generalization theory is committed to testing moral principles by asking what would happen if they were generally adopted among all persons at all times. A relativist generalization theory, on the other hand, will assess a moral principle for a subset of persons by asking what would happen if it were generally adopted by only

---

<sup>14</sup> The latter position is taken by Brandt (*A Theory*, *op cit.*, p. 287), Hooker (*Ideal Code*, *op cit.*, pp. 78–80), and Timothy Miller, “From Compliance, to Acceptance, to Teaching: On Relocating Rule Consequentialism’s Stipulations,” *Utilitas* xxxiii, 2 (2021): 204–20.

<sup>15</sup> See Parfit, *On What Matters*, *op cit.*, p. 356, 379; Scanlon, “Contractualism”, *op cit.*, p. 597; Brandt, *A Theory*, *op cit.*, pp. 212–213; and Hooker, *Ideal Code*, *op cit.*, p. 2, pp. 72–75.

that subset. Again, both kinds of generalization theory are defended in the literature: Parfit and Scanlon are universalists, while Brandt and Hooker are relativists.<sup>16</sup>

These two interpretive choices clearly affect the epistemic difficulty of generalization counterfactuals. Objectivism and universalism make the generalization theorist's epistemic task harder; subjectivism and relativism make it easier. I will start by laying out the more difficult epistemic problem faced by an *objective* and *universalist* generalization theory (sections II and III). I will then consider subjectivism and relativism as possible solutions to this problem (sections IV and V).

## II. OUR IGNORANCE OF UNIVERSALIST GENERALIZATION COUNTERFACTUALS

To assess a moral principle, a universalist generalization theorist must ask how the world would be if the principle were generally adopted across all persons—meaning in *all places and cultures* throughout *all of human history*. The next two sections argue that we are not in an epistemic position to answer this question: we *have no clue* what would happen if different moral principles were adopted by all of humanity.

Here are three reasons for pessimism about our ability to know what would happen if different moral principles were generally adopted. (Although I focus on universalist generalization counterfactuals, it is worth noting that these obstacles apply to relativists' smaller-scale generalization counterfactuals as well, just to a lesser degree.)

---

<sup>16</sup> See Parfit, *On What Matters*, *op cit.*, p. 377; Scanlon, *What We Owe*, *op cit.*, p. 328, 342; Brandt, *A Theory*, *op cit.*, p. 194, and "Fairness to Indirect Optimific Theories in Ethics," *Ethics* XCVIII, 2 (1988): 341–60, at pp. 350–351; and Hooker, *Ideal Code*, *op cit.*, p. 32. See also the references in footnote 34.

*II.1. Reason 1: Generalization counterfactuals are comparable in epistemic difficulty to political and evolutionary counterfactuals, about which we know very little.* The quickest route to epistemic pessimism about generalization counterfactuals is to notice how little we know about other counterfactuals that involve changes to human society.

First, consider *political counterfactuals*. If the United States implemented a universal basic income of \$20,000 per year, would this usher in a poverty-free golden age or lead to malaise and economic collapse? If the United States allowed unrestricted immigration, would that lead to increased economic prosperity and cultural tolerance, or strained social services and xenophobic backlash? These are devilishly complex empirical questions. They are not completely inscrutable—we can reasonably assign more confidence to some hypotheses than to others. But these counterfactual questions are sufficiently difficult that, even with lots of empirical data, the best answer we can manage is an educated guess.

Now compare a generalization counterfactual: what would happen if almost everyone accepted a moral principle requiring one to never kill under any circumstances, even in self-defense? How would the acceptance of this pacifism principle influence our systems of government, crime levels, or medical practices? This question is at least as difficult to answer as the above questions about universal basic income and immigration.

In fact, universalist generalization counterfactuals are far more difficult to assess than political counterfactuals, because they are much broader in scope. The political counterfactuals we just considered concerned the effects of legislation within a single country (the United States) at a single time (the near future). While when the universalist asks what the consequences of a principle's general adoption would be, they are asking what would happen if that principle were adopted by (a proportion of) all of humanity throughout history. We are counterfactually

legislating for all nations across all eras and trying to predict the result. Predicting the effects of principles' general adoption must be at least as hard as predicting the effects of legislation, and is likely to be much harder.

For a more sobering comparison, consider *evolutionary counterfactuals*: counterfactuals that suppose that humans had evolved different traits. A first example, mentioned in the introduction: what would happen if humans had never evolved religious belief? Would the absence of religion accelerate scientific progress by making people less willing to believe on the basis of faith? Or would science have never gotten off the ground without the scholarly institutions created by churches? A second: what would have happened if women had evolved to be as physically strong, on average, as men? Would this have prevented the development of patriarchy, leading to more just and enlightened societies? Or would greater physical strength have made women more inclined to war, resulting in a bloodier path for humanity?

These evolutionary counterfactuals share much in common with universalist generalization counterfactuals. Both involve imagining a substantial change in human dispositions across all of history and trying to predict the effects. Given the similarities in their contents, our epistemic position regarding evolutionary counterfactuals and universalist generalization counterfactuals should be roughly the same. This should worry universalist generalization theorists, as it seems clear that we have no clue how to answer the evolutionary counterfactual questions above. There is no way to tell whether science would have developed faster or slower without religion, or whether we would have been more or less warlike if women were as physically strong as men. I see no reason why our grip on universalist generalization counterfactuals should be any more secure.

*II.2. Reason 2: The causal mechanisms that mediate the effects of principle adoption are extremely complex.* To determine what the world would be like if a principle were generally adopted, we have to suppose that most people accept that principle, rerun history from this new starting point, and predict what will follow. Attending to the complexity of this task provides a more direct argument for epistemic pessimism.

One way to get a sense for this complexity is to attempt to list the components of human life that a principle's adoption might affect. To give a list that is certainly incomplete, the general adoption of any moral principle will influence a society's political institutions; military practices; punishment and law enforcement practices; legal practices; educational institutions and norms; scientific and technological development; medical practices; journalistic and press institutions; religious institutions and beliefs; artistic practices; sporting practices; diet and agriculture; racial norms and ideologies; gender norms and ideologies; family structures; parenting practices; formal and informal hierarchies of authority; sexual norms; divisions of labor; economic systems; conceptions of property rights; modes of transportation; migration patterns; and treatment of animals and the environment. To predict a principle's consequences, we would have to estimate its direct effects on at least these elements of human life.

That sounds laborious, but perhaps not undoable. Where the difficulty really ratchets up is further down the causal chain. For every direct effect a principle's adoption has on some element of society, that effect will itself have further effects on other elements of society. Those secondary effects will have their own tertiary effects, which will have quaternary effects, and so on, through countless orders of causal influence. There will also be feedback loops, where a principle's influence on one element of society sets off a chain of effects that comes back and further changes

that very element, which then sets off another chain that feeds back, and so on, amplifying the principle's effects potentially indefinitely.

To illustrate, consider the question: how would the general adoption of a principle forbidding lying under any circumstances influence the practice of marriage? That question is difficult enough—would a stronger norm against lying fortify the institution of marriage by reinforcing trust, or undermine it by forcing spouses to divulge relationship-destroying secrets? But suppose that we had an answer to this direct question. Then we would still have to ask: how would the resulting changes in marriage practice affect gender norms? And then: how would those changes in gender norms affect economic divisions of labor? How would those economic changes influence religious institutions? Finally, bringing us to the feedback loop: how would those religious changes affect the practice of marriage?<sup>17</sup>

To predict the consequences of adopting a moral principle, we have to predict the course of its reverberations through the enormous, interconnected causal system of human society. Attending to the requirements of this task should undermine our confidence in any intuitions we may have about what would result from a principle's general adoption.

*II.3. Reason 3: The methods philosophers use to assess generalization counterfactuals are unreliable.* Despite the considerations above, the fact remains that we *do* have intuitions about what would result from different principles' general adoption. Some generalization counterfactuals strike us immediately as likely to be true; others seem obviously far-fetched. Generalization theorists rely widely on these snap judgments. To determine whether their theory endorses a moral

---

<sup>17</sup> Compare Joseph Henrich's argument that the Catholic prohibition on cousin marriage, by pushing people to marry outside their families, initiated a cascade of effects on Western society that culminated in the industrial revolution (Joseph Henrich, *The WEIRDest People in the World* (New York: Farrar, Straus and Giroux, 2020)).

principle, most generalization theorists simply imagine a world where that principle is generally adopted and report their intuitive sense of what that world would be like. Brandt confidently endorses this methodology:

All the [rule consequentialist] moralist has to know ... [is] whether there would be more public benefit if such a rule were taught and widely prevalent, as compared with no rule at all or a somewhat different rule ... We hardly need a social science survey to determine this.<sup>18</sup>

Continuing, Brandt declares: “These problems seemingly can be resolved in the same way, just by taking thought!”<sup>19</sup>

I disagree. We have no reason to think that armchair intuition is reliable in this domain. One need not be skeptical about the philosophical use of intuitions in general to doubt their reliability here. Few maintain that intuition is a reliable guide to empirical questions such as who will win an election or whether there is life in Jupiter’s atmosphere. The question of what would result from a principle’s general adoption is similarly an empirical question. To put the point somewhat rhetorically: predicting a principle’s effects by consulting one’s intuitions about what would happen if it were generally adopted is like predicting whether a new drug will treat cancer by consulting one’s intuitions about what would happen if a patient took it.

To be fair, most people know more about human nature than they do about oncology. Intuition can be reliable on empirical matters when it is guided by sufficient background

---

<sup>18</sup> Richard Brandt, “Conscience (Rule) Utilitarianism and the Criminal Law,” *Law and Philosophy* XIV, 1 (1995): 65–89, at p. 77.

<sup>19</sup> *Ibid.*, p. 85.



knowledge. Perhaps our knowledge of human nature is enough to make our intuitions reliable in the sociological domain?

The data do not support this view. In what is likely the most thorough study of social prediction to date, the psychologist Philip Tetlock asked a group of 284 expert professionals in politics and economics to make hundreds of geopolitical predictions over the course of several years, and then waited to see which of their predictions came true.<sup>20</sup> Tetlock asked his subjects to make concrete, verifiable predictions: for example, he might ask them to predict whether the incumbent party will stay in power in the next US election, whether Japanese Gross Domestic Product growth will accelerate, decelerate, or remain constant over the next 5 years, or which countries will acquire the capacity to create weapons of mass destruction in the near future.<sup>21</sup> The results are unambiguously humbling. Tetlock puts it colorfully: “The average expert was roughly as accurate as a dart-throwing chimpanzee.”<sup>22</sup> The experts’ average performance was barely better than the “chimp” strategy of assigning equal probability to all available outcomes.<sup>23</sup>

As Tetlock is quick to caution, this average obscures important individual differences. Tetlock divides his subjects into two categories, “foxes” and “hedgehogs” (an homage to the aphorism “the fox knows many things, the hedgehog knows one big thing”). Tetlock’s hedgehogs made their predictions by applying grand theories of politics and human nature. The foxes made

---

<sup>20</sup> Philip Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton: Princeton University Press, 2005).

<sup>21</sup> For more detail, see *ibid.*, pp. 239–52.

<sup>22</sup> Philip Tetlock and Dan Gardner, *Superforecasting: The Art and Science of Prediction* (New York: Broadway Books, 2015), p. 68.

<sup>23</sup> Tetlock, *Expert Political Judgment*, *op cit.*, p. 51.

their predictions on a case-by-case basis and paid assiduous attention to empirical data. This empiricism is likely why the foxes vastly outperformed the hedgehogs, many of whom performed worse than chance.<sup>24</sup> These results suggest that the data-free methodology favored by philosophers is particularly unreliable.

But even the foxiest forecasters soon reach their limits. After publishing his depressing initial study, Tetlock dedicated his research to improving predictive accuracy, training a group of “superforecasters” whose predictions decisively beat chance. However, these superforecasters’ remarkable accuracy quickly declines at larger temporal horizons. Summing up decades of research, Tetlock concludes: “There is no evidence that geopolitical or economic forecasters can predict anything ten years out beyond the excruciatingly obvious.”<sup>25</sup> If the best forecasters’ reliability is limited to a ten-year horizon, it is hard to see how universalist generalization counterfactuals, which range across all of human history, could lie within our ken.

Taken together, the three considerations reviewed in this section make a compelling case for pessimism about our ability to know which universalist generalization counterfactuals are true. Determining the effects of a principle’s general adoption requires predicting the output of the staggeringly complex causal mechanisms of human society. When we attempt the similar tasks posed by political and evolutionary counterfactuals, we are quickly reduced to blind guessing. Even using all the empirical data available, our prospects for knowing the effects of principles’ general adoption are dim. And if we follow philosophical convention and assess generalization counterfactuals from the armchair, we should not expect to do any better than Tetlock’s dart-throwing chimpanzee.

---

<sup>24</sup> *Ibid.*, pp. 78–80.

<sup>25</sup> Tetlock and Gardner, *Superforecasting*, *op cit.*, pp. 243–44.

### III. WE HAVE NO CLUE WHAT UNIVERSALIST GENERALIZATION THEORIES IMPLY ABOUT WHAT IS

#### RIGHT AND WRONG

The arguments of the last section may not be enough to dispel some readers' intuitive sense that the particular counterfactual claims generalization theorists make must be correct. *Of course* the world would be terrible if people told lies willy-nilly, killed whenever they deemed it optimal, or broke promises that were inconvenient to keep!

I think that this aura of plausibility has been encouraged by a tendency to focus on all-or-nothing comparisons. Consider, for instance, Nicholas Southwood's contractualist defense of commonsense deontological constraints (which he calls "potential rights"):

Imagine a world in which potential rights were absent. In such a world it would be permissible to kill others for fun, rape them, torture them, bully them, steal their possessions, manipulate them, lie to them, break promises that one has made to them, and display utter disregard for their plight as they drown in a puddle by the side of the road. Such a world would be truly disastrous.<sup>26</sup>

I agree that a world in which people feel permitted to "kill others for fun, rape them, torture them," and so on would be a bad place to live. But granting that general adoption of the principles of common sense would be superior to complete moral anarchy, this is not enough to show that contractualism or rule consequentialism vindicate commonsense moral views. To show this, the generalization theorist must establish that general adoption of the principles of commonsense morality would be superior to the general adoption of *any* alternative set of principles.

---

<sup>26</sup> Southwood, *Contractualism, op cit.*, p. 161. To be fair, Southwood then admits "This is too quick" (p. 162). But his further arguments consider only one additional alternative principle: act consequentialism.

Consider the exception clauses attached to many commonsense moral principles. We think that it is wrong to kill—except in self-defense. We think that it is wrong to lie—except when throwing a surprise party. We think that we are obligated to keep our promises—unless the promise was coerced. To capture these nuances, the generalization theorist needs to compare two nearby possible worlds: say, a world in which it is generally accepted that coerced promises are binding versus a world in which it is generally accepted that coerced promises are invalid. When we compare these worlds, we cannot point to widespread murder, rape, and torture to differentiate one from the other. We need a finer-grained grasp of the effects of these principles’ general adoption—a grasp that I have argued we do not have.

To illustrate the point in more detail, I shall consider some moral principles in the domain of honesty that are counterintuitive without being anarchic. I will offer counterfactual hypotheses that, if true, would lead a universalist generalization theory to endorse the counterintuitive principles. Given the arguments for epistemic pessimism offered in section II, I submit that we have no reason to think that these alternative hypotheses are less likely to be true than the counterfactual hypothesis that supports the commonsense principle. This means that a universalist generalization theory is just as likely to endorse one of the counterintuitive principles as it is to endorse the commonsense principle. I conclude that our ignorance of generalization counterfactuals leaves us unable to know what universalist generalization theories imply about which actions are right and wrong.

What principle governing honesty would produce the best consequences if generally adopted?<sup>27</sup>

---

<sup>27</sup> For simplicity, I shall assume rule consequentialism, but I think that the counterfactuals below also support their corresponding principles on contractualism.

The commonsense moral principle can be roughly stated as follows:

COMMONSENSE HONESTY: It is wrong to lie, except in certain special contexts (surprise parties) and when lying is necessary to prevent severe harm (the murderer at the door). However, one is not obligated to volunteer all the information one has to others, even if they want to know it.

Here is a counterfactual argument in favor of this principle:

THE PRO-COMMONSENSE STORY: Testimony is an essential mechanism for sharing information, building relationships, and maintaining cooperation. We would not be able to reap these benefits unless we were able to trust each other's testimony, and we would not trust each other's testimony unless a strong prohibition on lying were generally accepted. However, we can allow that in certain rare situations it is permissible to lie without destabilizing the practice.

While sharing information is generally beneficial, a duty to offer others any information they want would be too burdensome. A society that accepted the commonsense principle, forbidding lying but allowing omission of information, would strike the best balance between facilitating trust and enabling privacy.<sup>28</sup>

This story is quite plausible, but it is not the only way things could go. Here are counterfactual stories in support of two less intuitive principles.

First, consider

EXTREME HONESTY: Lying is wrong no matter what. Moreover, it is always wrong to withhold information that you expect someone else wants to know.

---

<sup>28</sup> Compare Seana Shiffrin, *Speech Matters: On Lying, Morality, and the Law* (Princeton: Princeton University Press), pp. 21–26.

Here is why the general adoption of *Extreme Honesty* might be best:

THE PRO-EXTREME STORY: If *Extreme Honesty* were generally adopted, the glue of society would be so strong as to make great things possible. The transparency and full information necessary for economists' efficient markets would actually obtain. Demagogues would have no power because they would feel obligated to tell the complicated truth. Moral accountability would be more effective because people would admit their faults. Scientific progress would be supercharged by scientists' sense of obligation to reveal the flaws in their theories and experiments.

Sure, sometimes people would tell the truth even when it would lead to disaster, but those situations would be so rare that their costs would be far outweighed by the benefit of unshakable general trust in testimony. And yes, people would not have much privacy in this world—but since they were not raised to expect privacy, they would not mind.

Moving in the opposite direction, consider

WEAK HONESTY: Other things being equal, one ought to avoid lying. But if lying will benefit you or people you care about, it is permissible to do so.

Here is why generally adopting *Weak Honesty* might be best:

THE PRO-WEAK STORY: Although some amount of honesty is necessary to the functioning of society, an adequate dose of deception is just as important. The fact that people often lie is what led us to develop many of the cornerstones of civilization, from courts to newspapers to dual-account ledgers. If the *Commonsense Honesty* principle were widely internalized, we would never have established these mechanisms of accountability and enforcement. As a result, many

other beneficial features of society would never have come about, such as contracts, due process, and investigative journalism.

Lying is also instrumentally essential to various valuable human endeavors. Religion, a source of personal fulfillment and moral guidance, can be sustained only if its representatives need not always believe what they say. Political movements of the kind that have brought great moral progress, from democracy to women's suffrage to the end of slavery, could not have gained momentum if their leaders were unwilling to embellish the truth to inspire crowds.

Some might object that if the honesty principle were this weak, then no one would trust others' testimony, and the whole practice would dissipate. But this is incorrect: practices can continue to function even when violation of their norms is common. The ubiquity of divorce has not ended the practice of marriage. The practice of testimony would thus survive the general adoption of *Weak Honesty*.

When reading these counterfactual stories, it is tempting to consult one's intuitions about their plausibility. To some readers, *The Pro-Commonsense Story* may seem quite plausible, whereas *The Pro-Weak Story* and *The Pro-Extreme Story* may seem dubious. While these intuitive reactions are natural, they cannot be trusted. The lesson of Tetlock's data is that our gut instincts about the above stories' plausibility are not a good guide to the truth. The fact that one finds *The Pro-Commonsense Story* more intuitively plausible than *The Pro-Weak Story* or *The Pro-Extreme Story* is no evidence, or at best very weak evidence, that it is more likely to be true.

Our intuitions might be a decent guide to whether a particular mechanism of cost or benefit would be likely to occur. Our background knowledge of human nature might give us sufficient basis to say, for example, that increased lying would undermine trust in testimony to some extent.

But none of the effects described in the stories above seem intuitively crazy on their own. License to lie *would* help leaders to motivate crowds; extreme honesty *would* provide some benefit to science. To argue that *The Pro-Commonsense Story* is more likely to be true than *The Pro-Weak Story* or *The Pro-Extreme Story*, we would have to assess the relative magnitude, frequency, and downstream implications of these positive and negative effects. This is where intuition becomes unreliable, and empirical data are necessary.

One way to frame my argument is as a challenge. Suppose you wanted to argue that *The Pro-Commonsense Story* is more likely to be true than *The Pro-Weak Story* or *The Pro-Extreme Story*. What would you need to do to defend this claim? Compare a political counterfactual that economists have recently debated: if the US Federal Reserve raised interest rates by 0.25%, how much would that slow inflation, and how likely would it be to cause a recession? Suppose you entered this debate and argued that the benefits of slowing inflation would outweigh the risk of causing recession, and offered as your only evidence for this claim your intuition that this would be so. You would be laughed out of the room. How an interest rate hike will affect the economy is an empirical question that can only be answered with empirical evidence.

Suppose, then, that you set out to defend *The Pro-Commonsense Story* with empirical evidence. For instance, against *The Pro-Weak Story*, you might look for evidence that increased license to lie would destroy the practice of testimony. To do this, you would need to find data on how trust in testimony depends on the frequency of lying across different societies. If you were able to procure this evidence, you would then need to defend the claim that we can extrapolate from it to the effects of lying across all societies in human history. If you pulled off this Herculean task, however, you would not be done. You would then need to estimate the magnitude of all the



other potential effects of different honesty principles, and back up those estimates with empirical evidence. I doubt that the data required to meet this challenge even exist.

I submit that, at least without *much* more empirical evidence, we have no reason to believe that *The Pro-Commonsense Story* is more likely to be true than *The Pro-Weak Story* or *The Pro-Extreme Story*. Thus, we have no reason to think that a universalist generalization theory is more likely to endorse *Commonsense Honesty* than it is to endorse *Extreme Honesty* or *Weak Honesty*.

Our evidence says so little about what would follow from any principle's general adoption that we can generate a panoply of counterfactual hypotheses that are all equally compatible with it. The depth of our ignorance gives us proportionately wide latitude for speculation. So, for any question of moral importance, we can cook up counterfactual hypotheses that, if true, would lead any universalist generalization theory to endorse both intuitive and counterintuitive answers. Because we have no evidence that favors one hypothesis over the others, we have no reason to think that our generalization theory is more likely to endorse the intuitive principles than it is to endorse the counterintuitive ones. We thus have no clue what universalist generalization theories imply about which actions are morally right and wrong. For all we know, universalist contractualism or rule consequentialism might entail that it is permissible to lie whenever convenient, or that it is wrong to lie to the murderer at the door.

#### IV. TWO UNIVERSALIST REPLIES

The main response I shall consider to the cluelessness problem is to adopt moral relativism. But first it will be worthwhile to ask whether there is any way to make generalization theories more epistemically tractable while retaining universalism. I will discuss two proposals.

*IV.1. Reply 1: Shrink the testing population.* Universalist generalization counterfactuals are so inscrutable because they require us to predict moral principles' effects across all of human history. This suggests a solution: rather than asking what would happen if different principles were generally adopted by all of humanity, we might ask what would happen if principles were generally adopted by a more restricted population—say, by a single generation of a society of moderate size.

This is the motivating idea behind the relativist response, but our question here is whether it could be employed by the universalist. Could we test moral principles on a restricted population while also holding that the resulting principles apply to all persons? To do so, we would have to reject an assumption I call

INCLUSION: If a person is subject to a moral principle, then they (or a sufficiently similar counterpart) must be included in the population that a generalization theory uses to test that principle.

A theory that retains moral universalism while rejecting *Inclusion* would hold that the moral obligations of all persons depend on what would happen if different principles were adopted by one *special population*, a restricted set of people living in a particular place at a particular time. Such a theory will face three major problems. The first is a problem of *arbitrariness*. How do you select the special population? What reason could you give for saying that the special population should be the present United States, rather than Egypt in 5000 BCE? Any choice of the special population will be arbitrary, and so the resulting moral principles will be arbitrary as well. This leads to a problem of *authority*: why should the moral principles that are best in the circumstances of an arbitrarily chosen population be binding on those of us who live in different circumstances? Finally, there is the problem of *moral knowledge*: how can people outside the special population know what is right and wrong? If I do not know the circumstances of the special population, then

I cannot know what moral principles are vindicated by the true generalization theory. So, a theory that rejects *Inclusion* implies that anyone in this position will be excluded from moral knowledge. These costs strike me as too steep. If this is right, and rejecting *Inclusion* is untenable, then universalist generalization theories are committed to testing moral principles on all persons.

*IV.2. Reply 2: Adopt subjectivism.* Perhaps the epistemic problem can be avoided if we switch to a *subjective* generalization theory, which assesses moral principles based not on what would actually happen if they were generally adopted, but rather on the expected consequences of their general adoption given our evidence. Brad Hooker takes this route, formulating his version of rule consequentialism to say that the true moral principles are those the general adoption of which would produce the greatest expected value.<sup>29</sup>

The key question for a subjective generalization theory is: what body of evidence do we use to assess the expected consequences of principles' general adoption?<sup>30</sup>

A universalist must hold that there is a single body of evidence that determines the moral principles that apply to all persons, because they take there to be a single set of moral principles that apply to all persons. They cannot hold, for instance, that I ought to follow the moral principles with the highest expected value *given my evidence*, or *given the publicly available evidence in my society*.<sup>31</sup> For this leads to relativism at the level of individuals or societies.

The question becomes: what is the relevant body of evidence? Here the subjective universalist faces a dilemma. Either the relevant body of evidence includes information available to only some people, or it does not.

---

<sup>29</sup> Hooker, *Ideal Code*, *op cit.*, p. 2, pp. 72–75. Hooker's theory is also relativist, but we can set that aside for now.

<sup>30</sup> As far as I can tell, Hooker does not discuss this question.

<sup>31</sup> See Brandt, *A Theory*, *op cit.*, p. 13.

Suppose we opt for the former and say that the relevant evidence contains information that is accessible only to some restricted population. Then we shall face the same problems that result from rejecting *Inclusion*. *Arbitrariness*: any choice of which population's body of evidence is relevant will be arbitrary, and so the resulting moral principles will be arbitrary. *Authority*: why should I be bound by moral principles based on evidence unavailable to me? *Knowledge*: anyone who lacks access to the relevant body of evidence will be unable to attain moral knowledge.

To avoid these problems, we must hold that the relevant body of evidence contains only information available to all persons. But this body of evidence will have to be extremely minimal: consider what evidence is shared in common by a citizen of Zhou dynasty China, a pre-colonization Aztec, and a person alive in the year 2500. This leads to a problem of indeterminacy. In order for a subjective generalization theory to endorse a determinate set of moral principles, there must be a single set of moral principles that has the highest expected value given the relevant body of evidence. But a body of evidence so minimal as to be shared by all persons will have almost nothing to say about a topic as epistemically opaque as universalist generalization counterfactuals. It is hard to see how this minimal evidence could favor a determinate set of moral principles. More likely, there will be a wide range of sets of moral principles tied for best, simply because the relevant body of evidence gives us no basis for distinguishing between them.<sup>32</sup> We thus convert ignorance into indeterminacy, trading the epistemic problem of our not knowing what

---

<sup>32</sup> Hooker proposes a tie-breaker clause that tells us to choose, of the tied-for-best options, the set of principles closest to the conventional morality of our society (*Ideal Code, op cit.*, p. 114). The trouble is that, given how many principles will be tied for best, all the work of principle selection will be done by the tie-breaker clause, resulting in the clearly false view that we are always morally required to obey conventional morality.

moral principles are true for the metaphysical problem of there being no fact of the matter as to which principles are true, as no set is most favored by the evidence available to all.

Any subjectivist generalization theory, even a relativist one, must be wary of this indeterminacy problem. Whatever evidence the subjectivist uses, if that evidence does not give us any clue about the objective effects of a principle's general adoption, it will not favor a determinate set of moral principles. I think that we are in this position already, with all the data at our disposal. I argued in section III that our evidence gives us no reason to favor *Commonsense Honesty* over *Extreme Honesty* or *Weak Honesty*. If we are in this position, that means not just that we do not know which principle an objectivist generalization theory endorses, but also that it is indeterminate which principle is endorsed by a subjectivist generalization theory given our evidence. The upshot is that even subjectivists need to show that we are in a decently good epistemic position to assess the *objective* effects of principles' general adoption—otherwise, indeterminacy looms.

The problems with both the universalist solutions considered in this section point in the direction of relativism. If we wish to test moral principles on smaller populations while retaining *Inclusion*, we must take different principles to apply to different populations. And if we wish to test moral principles using a substantive body of evidence that is available to all persons subject to the principles, then we should take different moral principles to apply to groups with different bodies of evidence. The most promising way for a generalization theorist to avoid cluelessness is thus to adopt moral relativism.

#### V. TO AVOID CLUELESSNESS, RELATIVIST GENERALIZATION THEORIES MUST BE EXTREME

I have argued that any generalization theory that accepts moral universalism faces a problem of cluelessness. If a generalization theorist adopts moral relativism, holding that different basic moral

principles apply to different groups of persons, their epistemic task becomes more tractable, for two reasons.<sup>33</sup> First, they can test moral principles on a smaller scale. It is easier to assess what would happen if different principles were adopted by a spatiotemporally restricted group than what would happen if they were adopted by everyone throughout history. Second, once we adopt relativism, subjectivism becomes more tenable. The relativist can say that the moral principles that apply to a given group are those that best satisfy their theory's principle selection criteria relative to the evidence available to that group. This evidence will contain much more information than the evidence shared by all persons, and so will be more likely to favor a determinate set of principles.

Adopting relativism may seem a congenial solution, as many generalization theorists—especially rule consequentialists—are already relativists.<sup>34</sup> It is worth noting, however, that two of

---

<sup>33</sup> Moral relativism claims that different *basic* or *nonderivative* moral principles apply to different groups of people. This is different from the weaker claim, compatible with universalism, that the same basic principle can endorse different kinds of actions in different circumstances. If it is rude to belch at dinner in one culture but rude not to belch in another, one should belch accordingly. This does not show that different moral principles apply in different cultures: it shows that the same basic principle (say, “avoid rudeness”) recommends different actions in different contexts. Compare Scanlon, *What We Owe*, *op cit.*, p. 329.

<sup>34</sup> Richard Brandt, Leonard Kahn, and David Copp relativize to societies: see Brandt, *A Theory*, *op cit.*, p. 194; Leonard Kahn, “Rule Consequentialism and Scope,” *Ethical Theory & Moral Practice* xv (2012): 631–46; and David Copp, *Morality, Normativity, and Society* (Oxford: Oxford University Press, 1995), p. 218. Ryan Jenkins and Tim Mulgan relativize to groups within societies: see Ryan Jenkins, “Rule Consequentialism and Moral Relativism,” *Journal of Philosophical Research* xli (2016): 527–37; Tim Mulgan, “One False Virtue of Rule Consequentialism, and One New Vice,” *Pacific Philosophical Quarterly* 77 LXXVII (1996): 362–73, at pp. 367–71. Brad Hooker, Dale Miller, and Nicholas Southwood relativize to generations or times: see Hooker, *Ideal Code*, *op cit.*, p. 32; Dale Miller, “Moral Education and Rule Consequentialism,” *The Philosophical Quarterly* LXXI (2021): 120–40, at p. 135; Southwood,

the most famous generalization theorists, Derek Parfit and T. M. Scanlon, explicitly reject relativism.<sup>35</sup>

If the arguments so far succeed, however, a generalization theorist *must* be a relativist to avoid cluelessness. This is an important result. The idea that contractualism and rule consequentialism are pressured toward relativism is not widely accepted. The word “relativism” does not occur in the *Stanford Encyclopedia of Philosophy* (SEP) article on contractualism,<sup>36</sup> and at the time I am writing, a search for “relativism” in the Philpapers.org category on contractualism yields 2 articles out of 286.<sup>37</sup> The SEP article on rule consequentialism does acknowledge that many rule consequentialists are relativists,<sup>38</sup> but Philpapers does no better: “relativism” presently yields 2 papers out of 154 in the rule consequentialism category.<sup>39</sup>

If it were widely recognized that contractualism and rule consequentialism are committed to moral relativism, the landscape of the debate between the Big Theories in normative ethics would look very different. These theories would be divided into two major camps: a universalist

---

*Contractualism, op cit.*, p. 116. I am grateful to a reviewer for suggesting that I discuss relativism at more length, as well as for suggesting several of these references.

<sup>35</sup> Parfit, *On What Matters, op cit.*, p. 382; Scanlon, *What We Owe, op cit.*, p. 328, 342, 348.

<sup>36</sup> Elizabeth Ashford and Tim Mulgan, “Contractualism,” in Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition).

<sup>37</sup> See <https://philpapers.org/browse/moral-contractualism?catq=relativism&sort=relevance&uncat=&setAside=&Search=Search>

<sup>38</sup> Brad Hooker, “Rule Consequentialism,” in Edward Zalta and Uri Nodelman (eds.) *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition).

<sup>39</sup> See <https://philpapers.org/browse/act--and-rule-consequentialism?catq=relativism&sort=relevance&uncat=&setAside=&Search=Search>

camp including act consequentialism, Kantian deontology, Rossian pluralist deontology, and virtue ethics, and a relativist camp including contractualism and rule consequentialism.<sup>40</sup> Which side of this theoretical menu one selects from would then depend on how one sides in the debate between universalism and relativism. In short, even if the relativist solution to the cluelessness problem works perfectly, this paper has already defended a surprising conclusion: to be viable, generalization theories must accept moral relativism.

The argument does not stop there, however. I shall now contend that by accepting relativism, generalization theorists merely trade one set of problems for another. In addition to the objections that all relativists face (section V.1), relativist generalization theories are under pressure to make their testing groups quite small, resulting in an extreme form of relativism (section V.2).

*V.1. Problems with Relativism in General.* The simplest reason to reject a relativist generalization theory is that moral relativism is false. I am persuaded of this, but some of my interlocutors may not be. But we can all agree that if they wish to avail themselves of the relativist solution to cluelessness, the generalization theorist must respond to the considerations that motivate many philosophers to accept universalism. Although Hooker's theory is temporally relativist, he expresses the intuition behind moral universalism eloquently:

To go down [the relativist] road is to turn our backs on one of the traditional attractions of rule-consequentialism—namely, its basis in the idea that morality

---

<sup>40</sup> Although if Kant's Formula of Universal Law is a generalization theory, it will be forced into the relativist camp as well (to Kant's posthumous chagrin).



should be thought of as a collective, shared code. As Gert ... writes, "...The requirements of morality apply to all rational persons."<sup>41</sup>

The compelling thought that morality is *public* seems to count in favor of universalism. One might think that morality provides standards to which we can hold each other accountable, which requires those standards to be shared. One might also be motivated to accept universalism simply by first-order intuitive judgments: one might think that it is (other things being equal) wrong for *anyone* to lie, or to kill, or to break a promise, wherever and whenever they are; and importantly, it is wrong *for the same reasons* it would be wrong for someone in one's own time and place to do so.

The intuitive plausibility of universalism undercuts one of the major arguments in favor of generalization theories: that they vindicate and explain our commonsense moral judgments. Hooker defends rule consequentialism on this basis: "The best argument for rule-consequentialism is that it does a better job than its rivals of matching and tying together our moral convictions."<sup>42</sup> The arguments of this paper make this defense less persuasive. If a generalization theorist sticks to universalism, they have no basis for the claim that their theory matches our moral convictions, because they have no clue what moral verdicts their theory implies. While if the generalization theorist appeals to relativism to avoid cluelessness, they thereby deny a moral conviction many of us (Hooker included) find compelling: that the same basic moral principles apply to all persons.

The second task that the generalization theorist takes on by adopting relativism is to specify how to individuate the groups to which moral principles are relative. Let us call a group of persons who share the same set of basic moral principles a *moral cohort*. Universalism says that there is

---

<sup>41</sup> Hooker, *Ideal Code, op cit.*, pp. 87–88, citing Bernard Gert, *Morality: Its Nature and Justification* (New York: Oxford University Press, 1998), p. 216.

<sup>42</sup> *Ibid.*, p. 101.

only one moral cohort—everybody—while relativism says there are multiple distinct moral cohorts. The question for the relativist is: how do we distinguish one moral cohort from another?

The problem is that there seems to be no non-arbitrary way of sorting people into moral cohorts. The relativist generalization theorists cited above tend to identify their moral cohorts with “societies” or “generations.” But these are vague terms, raising the worry that any way of making them determinate will be arbitrary. People are born continuously through time, so what basis could we have for, say, counting people born from 1980–90 as a single generation, rather than 1985–95 or 1982–92? The difficulty of distinguishing societies is especially vivid in today’s globalized world, where any person can potentially interact with virtually any other. Imagine a woman who was raised in Milwaukee, went to university in London, splits her time between Hong Kong and New York for work, and married a man from Toronto with family in India. Which “society” is she a member of? The worry here is not that relativists have failed to answer a fiddly question of formulation. The worry is about arbitrariness. If the relativist’s method of distinguishing moral cohorts is arbitrary or vague, then their theory’s moral verdicts will be arbitrary or vague.

Perhaps because they recognize the difficulty of the task, relativist generalization theorists have tended to avoid discussing how to distinguish moral cohorts. Southwood says in a footnote: “This raises the question of how we differentiate generations. I shall not seek to answer this question here.”<sup>43</sup> Regarding how to distinguish societies, Brandt writes “I propose not to worry about this question.”<sup>44</sup> Relativists cannot brush aside this question so casually: the viability of their theory depends on whether they can give a principled answer.

---

<sup>43</sup> Southwood, *Contractualism, op cit.*, p. 116.

<sup>44</sup> Brandt, “Fairness,” *op cit.*, p. 351.

*V.2. Epistemic pressure to shrink moral cohorts.* Any relativist faces the daunting task of individuating moral cohorts. But a relativist generalization theorist has a further problem: to avoid cluelessness, they need their moral cohorts to be epistemically tractable. The question becomes: how small do our moral cohorts need to be, if they are to be epistemically tractable?

Let us look at the moral cohorts employed by extant generalization theorists. Start with Hooker, whose theory tells us to obey “the code of rules whose internalization by the overwhelming majority of everyone everywhere in each new generation has maximum expected value.”<sup>45</sup> Despite Hooker’s insistence that “I favour one code applying to everyone,”<sup>46</sup> this view is quite radically relativist. As Hooker admits, it implies that the basic moral principles can change from second to second, as “the code whose internalization has the highest expected value might change with each new day (indeed, minute!).”<sup>47</sup>

Hooker’s view is still not relativist enough to avoid cluelessness, however. To determine whether a moral principle holds for us at present, we have to ask what would happen if that principle were generally adopted by everyone from now until the end of human history. This is barely easier than the universalist’s question. If we have no clue what would happen if moral principles were adopted by everyone throughout time, we do not have much more of a clue what would happen if they were adopted by everyone from now on. Hooker’s view does not give us what we need to avoid cluelessness, which is to substantially shrink the groups on which we test moral principles. And Hooker’s appeal to expected value does not help either, for the reason

---

<sup>45</sup> Hooker, *Ideal Code*, *op cit.*, p. 32.

<sup>46</sup> *Ibid.*, p. 174.

<sup>47</sup> Brad Hooker, “Reply to Arneson and McIntyre,” *Philosophical Issues* xv (2005): 264–281, at p. 274.

mentioned in section IV: if we have no clue what will actually result from a principle's general adoption, our evidence will not favor a determinate set of moral principles.

How temporally constrained do our moral cohorts need to be to avoid cluelessness? If we look at the data, the answer appears to be no more than a decade: recall Tetlock's admonition that "there is no evidence that geopolitical or economic forecasters can predict anything ten years out beyond the excruciatingly obvious."<sup>48</sup> This suggests that an epistemically tractable generalization theory must take different basic moral principles to apply to people in different decades.

The implications of this view are extreme. It means that I am subject to different basic moral principles now than I was when I graduated from college. This raises tricky questions about cross-temporal accountability. Suppose I did something in college that was wrong by the principles that applied then, but is permissible by the principles that apply now. Should I apologize?

I doubt, though, that temporal relativism alone will be sufficient. To make our moral cohorts epistemically manageable, we will also need to relativize synchronically. Tetlock is less explicit here, but his superforecasters' successful predictions are usually spatially as well as temporally restricted. They are better than chance on questions such as "whether the United States and Afghanistan [will] reach an agreement on the continued presence of American troops."<sup>49</sup> This is far more localized than, "What would happen if everyone globally adopted a principle forbidding lying under any circumstances for the next decade?"

It is hard to say precisely how finely we need to divide moral cohorts synchronically in order to make reliable predictions (it is an empirical question!). My armchair guess is that our cohorts will need to be no bigger than countries, and likely smaller than giant countries such as

---

<sup>48</sup> Tetlock and Gardner, *Superforecasting*, *op cit.*, pp. 243–244.

<sup>49</sup> *Ibid.*, pp. 166–167.

China and India. If this is right, our moral cohorts will be smaller than what Brandt or Copp call “societies.” Whatever the details turn out to be, the conclusion that we need to relativize synchronically as well as diachronically is enough to raise problems. What moral principles apply at the UN General Assembly? At a wedding between a Brazilian and a Ugandan?

Relativists may object that these problems will not arise in practice, since the principles their theory favors will not differ much across different moral cohorts.<sup>50</sup> But we cannot assume this without argument. As I have repeatedly emphasized, the question of what would happen if a principle were generally adopted by a moral cohort is a difficult empirical question, even when asked on a smaller scale. We cannot simply assert that the principles that our generalization theory endorses for present Ugandans will be the same as those it endorses for present Brazilians. This claim must be defended with empirical argument. Show me the data!

Another reply appeals to a view that Hooker calls “incrementalism.”<sup>51</sup> In response to epistemic worries raised by Alison McIntyre, Hooker writes that “I retreat to the view that the currently accepted moral code should be revised if and only if such revisions have greater expected value than sticking with the status quo.”<sup>52</sup> The idea is that we start with the currently accepted moral code, and ask whether any piecemeal revisions to that code have a higher expected value than the status quo.<sup>53</sup> This task *might* be less difficult—although judging which revisions will increase expected value still requires our generalization counterfactuals to be epistemically tractable, and so the incrementalist will also be under pressure to relativize to small groups.

---

<sup>50</sup> Copp, *Morality*, *op cit.*, p. 223, and Brandt, “Fairness,” *op cit.*, p. 351 both argue along these lines.

<sup>51</sup> Thanks to a reviewer for suggesting that I address incrementalism.

<sup>52</sup> Hooker, “Reply,” *op cit.*, p. 277. For McIntyre’s objection, see “The Perils of Holism,” *op cit.*, pp. 255–260.

<sup>53</sup> Brandt endorses a similar idea in “Fairness,” *op cit.*, p. 350.

But there is a deeper problem with incrementalism, which McIntyre puts well: “[incrementalism] involves abandoning the aim of explaining, justifying or revising any of our actual moral convictions by reference to the content of an ideal moral code.”<sup>54</sup> Most normative ethicists are looking for a theory that answers the question of which actions are morally right and wrong, and why. Incrementalism does not attempt to answer that question. Instead, it switches from the question of which actions are right and wrong to that of which actions *we should teach people* are morally right and wrong. “Seek incremental improvements on the existing code” is a reasonable practical suggestion for moral teachers or legislators, but it is not a moral theory.

To summarize: to avoid cluelessness, it is not sufficient for a generalization theorist to adopt relativism. They must relativize to moral cohorts small enough for us to make at least somewhat reliable predictions about what would happen if those cohorts generally adopted different principles. I have argued that, to be epistemically tractable, a moral cohort should be roughly the size of a small nation over no more than a decade span. This means that the relativism a generalization theorist must accept to avoid cluelessness is extreme. It implies that the basic moral principles that apply to persons vary within a single lifetime and a single continent. This is not only counterintuitive but also raises thorny questions about how to distinguish moral cohorts non-arbitrarily and how to understand cross-cohort accountability.

Any theory that takes the truth of moral principles to depend on what would happen if they were generally adopted, as contractualism and rule consequentialism do, thus faces a choice between two unpalatable options. If they endorse moral universalism, they must test moral principles by asking what would happen if they were adopted by (a proportion of) all persons across space and time. Since we have no clue how to answer this question, we have no clue what

---

<sup>54</sup> McIntyre, “Perils,” *op cit.*, p. 260.

principles these theories endorse. If a generalization theorist tries to avoid cluelessness by adopting moral relativism and testing principles on smaller moral cohorts, they are pushed to the counterintuitive conclusion that the basic principles of morality vary between decades and nations. Contractualists and rule consequentialists are thus forced to choose between cluelessness and extreme relativism. I find neither of these options acceptable. I conclude that we should reject any moral theory that entails that what morality requires depends on what would happen if different moral principles were generally adopted.

#### VI. A THIRD OPTION FOR CONTRACTUALISTS

That last sentence is meant precisely: my thesis is that we should reject any moral theory that entails that what morality requires depends on what would happen if different moral principles were generally adopted. This only means that we should reject contractualism and rule consequentialism if they have this implication. This suggests an escape route: if one could reformulate contractualism or rule consequentialism so that they do not make the requirements of morality depend on generalization counterfactuals, the resulting theories would not be subject to my objection.

I doubt that this option is open to rule consequentialists. I see no way of interpreting rule consequentialism's principle selection criterion—the principles that *would* produce the best consequences *if* generally adopted—such that it does not appeal to generalization counterfactuals.<sup>55</sup>

---

<sup>55</sup> That said, Caleb Perl has recently proposed a version of rule consequentialism that does not invoke counterfactuals (Caleb Perl, "Solving the Ideal Worlds Problem," *Ethics* CXXXII, 1 (2021): 89–126). On Perl's account, the consequences of a rule are the actual effects that have been and will be caused by actions that rule classifies as morally

Contractualists can purge generalization counterfactuals from their theory, however. To be clear, contractualism as it is understood in the present literature *does* appeal to generalization counterfactuals. Contractualists widely hold that whether a principle can be reasonably rejected depends on what would happen if it were generally adopted, and routinely employ generalization counterfactuals in their moral explanations.<sup>56</sup>

To escape this paper's dilemma, contractualists need to say that whether a principle can be reasonably rejected does not depend on what would happen if it were generally adopted. This is perfectly coherent; nothing in the idea of reasonable rejectability entails that reasons for rejecting a principle must appeal to the effects of its general adoption. Instead, the non-generalization contractualist can say that reasons for rejecting principles depend solely on the principles' content, considered independently of the effects of their general adoption.<sup>57</sup>

Without appeal to the effects of principle adoption, however, the contractualist's resources for explaining moral principles are severely curtailed. They cannot assess moral principles by, as Scanlon suggests, "comparing the generic reasons for having [the principles'] protections with

---

right (p. 99). But the cluelessness problem seems to apply to this view too: are we in an epistemic position to assess what effects have been and will be caused by compliance with a moral principle throughout all human history?

<sup>56</sup> See the references in footnotes 11 and 12.

<sup>57</sup> A different route would be to shift from *rule contractualism* to *act contractualism*, which evaluates each action in terms of whether the decision to perform *that action* could be reasonably rejected (see Shelly Kagan, *Normative Ethics* (Boulder: Westview Press, 1998), pp. 242–243; Southwood, *Contractualism*, *op cit.*, p. 102; Hanoch Sheinman, "Act and Principle Contractualism," *Utilitas* XXIII, 3 (2011), 288–315; and Léa Bourguignon, "On the Possibility of Act Contractualism," *Australasian Journal of Philosophy*, forthcoming). Act contractualism merits more exploration, but I suspect that it will have a similarly difficult time offering non-circular moral explanations.



generic reasons for being free of the burdens imposed by the principles that provide them.”<sup>58</sup> For these are the protections and burdens that people *would* have *if* the relevant principles were generally adopted. For example, to explain why it is reasonable to reject a principle permitting promise-breaking, the contractualist cannot say that adopting such a principle would prevent us from enjoying the benefits of assurance and cooperation that promises provide.<sup>59</sup> For this is just the sort of plausible-sounding but epistemically dubious counterfactual that generated our problem in the first place. Instead, the non-generalization contractualist must appeal only to the principle’s content, saying that it is reasonable to reject a principle permitting promise-breaking because permitting promise-breaking is objectionable in itself. This explanation is circular: the moral judgment it yields as output is already contained in the standards of reasonable rejectability it takes as input.<sup>60</sup> But if the contractualist resolutely avoids appeal to generalization counterfactuals, all of their explanations are likely to take this unsatisfying form.<sup>61</sup>

## VII. CONCLUSION

---

<sup>58</sup> Scanlon, *What We Owe*, *op cit.* p. 223.

<sup>59</sup> See Scanlon, *What We Owe*, *op cit.*, Ch. 7.

<sup>60</sup> Others have charged that contractualism’s explanations are circular: see Philip Pettit, *The Common Mind* (New York: Oxford University Press, 1993), p. 302; and Michael Ridge, “Saving Scanlon: Contractualism and Agent-Relativity,” *Journal of Political Philosophy* IX, 4 (2001): 472–481.

<sup>61</sup> Liam Murphy also argues that contractualists should not appeal to the effects of principle adoption, but expresses optimism that the resulting non-generalization contractualism can be explanatory (Liam Murphy, “Nonlegislative Justification,” in Jeff McMahan et al. (eds.) *Principles and Persons: The Legacy of Derek Parfit* (Oxford: Oxford University Press, 2021), 247–274, at p. 255). It remains to be seen how these explanations will work.

I have argued that what is morally right and wrong cannot depend on what would happen if different moral principles were generally adopted. Any theory that takes the truth of a moral principle to depend on the effects of its general adoption is forced to accept one of two unacceptable conclusions: either that we have no clue what morality requires, or that the basic principles of morality vary from one nation to another and from one decade to the next. As contractualism and rule consequentialism both, on their traditional formulations, imply that the truth of a moral principle depends on the effects of its general adoption, this argument gives us reason to reject both theories. Contractualists can avoid this objection by saying that whether a principle is reasonably rejectable does not depend on the effects of its general adoption; however, this revision undermines contractualism's explanatory power.

If I could choose one lesson for readers to take away from this paper, it would be this: *the question of what would happen if a principle were generally adopted is an empirical question and should be treated as such*. If, despite this paper's arguments, some ethicists wish to hold on to a generalization theory, they must at least change their moral methodology. We cannot determine what would happen if a principle were generally adopted from the armchair; this question can be answered only with empirical evidence. If contractualists and rule consequentialists internalized this fact, their arguments would look very different. Neither Scanlon's *What We Owe to Each Other* nor Parfit's *On What Matters* cite a single empirical study. Hooker's *Ideal Code, Real World* cites two, discussing each for three sentences.<sup>62</sup> Normative ethics must either become much more empirical, or abandon the idea that the truth of a moral principle depends on what would happen if it were generally adopted.

---

<sup>62</sup> See Hooker, *Ideal Code, op. cit.*, p. 163 and p. 187. One exception to this trend is Miller, "The Rule-Consequentialist Response," *op. cit.*, who cites empirical data to defend rule consequentialist principles regarding climate change.