

**Let's Be Flexible: Our Interpretive/Explanatory Toolbox,
or In Praise of Using a Range of Tools.ⁱ**

David Henderson

Abstract:

This paper explores the role and limits of cognitive simulation in understanding or explaining others. In simulation, one puts one's own cognitive processes to work on pretend input similar to that one supposed that the other plausibly had. Such a process is highly useful. However, it is also limited in important ways. Several limitations fall out from the various forms of cognitive diversity. Some of this diversity results from cultural differences, or from differences in individuals' cognitive biographies. Such diversity is clearly important in history. Some sorts of such diversity are discussed, with attention to the results of contemporary cognitive science. It is argued that one must sometimes employ mixed (simulation-based/theory-based) strategies, and that sometimes what is done will be neither purely simulation nor purely theory-based.

Keywords: simulation, understanding, explanation, interpretation, epistemology

1. Introduction

Over the last few decades, philosophers such as Robert Gordon, Alvin Goldman, Shaun Nichols and Stephan Stich, and Karsten Stueber have developed understandings of simulation as a way of explaining others. It is said that, one uses one's own processes, at work on pretend input, to simulate others (rather as one might use a model in a wind tunnel to simulate what transpired in an episode). In using oneself as a model, one relies on the plausible similarities across human cognitive processes (but, in many cases, one need not be able to characterize those similarities in significant detail). Rather than applying a propositional understanding of the cognitive tendencies of human agents—applying theory—one simulates one's subjects by putting one's own cognitive processes to work on a range of pretend-beliefs and pretend-desires. It is thought that, such simulation is crucial to

one's ability to understand and explain the actions of others—that it is important for one's everyday facility in explaining others actions, for historical reconstructions, and for the human sciences generally. I believe that these claims are largely and importantly correct—that simulation is an important, even central, component within our cognitive toolbox.

It is common to think of simulation as having three phases: (1) a set-up phase, or *matching phase*, in which one assembles a set of pretend inputs (pretend-beliefs, pretend-desires, pretend-perceptions in imagined settings) that would plausibly be needed to simulate one's subject, (2) the *simulation phase* itself in which one's processes are deployed offline, and (3) an *attribution phase* in which one attributes dependencies in the subject's thinking or acting that parallel what obtained in the simulation. It is widely acknowledged that propositional information can inform simulation in several respects. First, it can inform the matching phase. For example, one may have information regarding the others' setting and the salient objects there experientially confronted, or regarding the beliefs widely held (or not held) in their community. One may draw on similar kinds of information regarding likely desires. Such information provides a basis for assembling in the matching phase a set of pretend-beliefs and pretend-desires to feed into one's own cognitive processes. It may indicate that some of one's own actual beliefs and desires were not likely held by one's subjects, so that those beliefs and desires must be “quarantined” from one's cognitive processes.ⁱⁱ Second, it seems that propositional information about cognitive similarities, and cognitive variations within populations and across populations could inform one's confidence in one's own cognitive similarity to one's subject. Thus, propositional information can inform the attribution phase in which one moves from undergoing a bit of processing to attributing to the subject parallel dependencies within the subject's processes. In short, it is commonly allowed that “theory” can inform simulation in the matching and attribution phases.

Still, perhaps as a result of the felt need draw some bright line to keep simulation clearly distinct from the application of theory, its proponents have commonly felt the need to keep theory—or propositional knowledge—out of the picture in that phase that is most distinctive of simulation—the simulation phase itself. I want to consider the prospects for such *simulative purity*—because I sense that the proponents are trying too hard. I want to suggest that we confront a continuum of cases in which simulation grades off into the application of theory.

My starting point here will be Karsten Stueber's, *Rediscovering Empathy*.ⁱⁱⁱ For my purposes here, this work is important as an example of the purity approach to simulation. In the next section, I explain how Stueber draws his bright line—insisting on a kind of purity of simulation. In what then follows, I indicate why I think that this line must commonly be transgressed. My purpose is not to subsume simulation to the application of theory—that would be a mistake. Instead, I believe that we humans are capable of using a fluid and nuanced combination of simulation and application of propositional information when understanding others—and that we commonly need to, and do, use these mixed epistemologies.^{iv}

2. The Quest for Purity, and the Idea of Simulation as Relevantly Uncontaminated

The central idea advanced by simulation theorists is that much of our ability to understand and explain the actions of others turns upon our ability to put our own cognitive processes to work on what we imagine, or conjecture, to be the beliefs, desires, and input with which our subjects worked in the episodes in question. In doing this, we need not apply or even have some informative propositional description of how we ourselves reason—nor need we antecedently have some informative account of the cognitive processes to which our subject is disposed. We can manage to explain when possessing no “theoretical” description of the processes in play in us, and no “theoretical” understanding of the processes in play in our subject. Instead of an “information-rich” grasp of what is going on, we employ an “information-poor” strategy in which we rely on processes in us to instance and thus exhibit transitions between contents that parallel those in our subject—our processes at work on those contents exhibit “where they lead” (at least in agents similar to ourselves). The whole has something of a hypothetical character: if the cognition occurring in us plausibly yields the beliefs, desires, and decisions that obtained in the episode in question—the explananda—then we have some basis for thinking the input from which we worked (the pretend-beliefs and pretend-desires, the imagined input from salient aspects the agent's environment, and so on) conditioned very similar processes in the agent—issuing in what then counts as explained. If the pretend-beliefs and pretend-desires feature in a wide range of apparently successful simulations—making sense of much else that our subjects apparently do, our confidence is further enhanced. If, on the other hand,

the simulation does not yield the beliefs, desires, or actions that are in effect our (provisional) explananda, we try again.^v If we persistently fail to simulate the agent, this may lead us to question whether the agent really thought or intended in any of the ways we had supposed initially. In other words, the simulation approach fits neatly into an inference to the best explanation framework. We may weigh competing simulations in terms of the range of putative thoughts and actions they allow one to explain—and perhaps for something like simplicity.

To the extent that simulation is to have wide and successful application, “nature [must have] provided all [or most] human beings with a mind that is structured and that functions psychologically in a similar manner.”^{vi} With significant qualifications, some of which are important for the discussion to follow, it has. Further, it seems that we are endowed with a facility for imagination—we can “use our cognitive machinery... to think about actual or possible worlds, to process beliefs or mere ‘make-beliefs’.”^{vii} This imaginative facility is significant in several respects. We use this ability when we plan—as we imagine what would happen were such-and-such to occur. We also use it when we “take up another’s point of view”—imagining how we would think, feel, and act were we in their subjective situation.

Let us look closer. We can begin with the three phases:

- (1) The Matching Phase in which one assembles a set of pretend-beliefs and pretend-desires on which to deploy one’s cognitive processes. One might understand pretend-beliefs and pretend-desires as beliefs that one adds to one’s own set of standing beliefs and desires. As almost never will what is added in this phrase be all the content that impacts one’s processes when one simulates, the default is that one’s own standing beliefs and desires are also employed in simulation. But, one will commonly need to prevent certain beliefs or desires that as a matter of fact one possesses from having a role in the processing to come—as the subject of one’s simulation may lack that information or those desires. Thus, the resources on which one’s processes will be deployed in the simulation phase to follow will be some set of beliefs and desires that is the result of both supplementing one’s own actual beliefs and desires and of quarantining off some of one’s beliefs and desires. In cases in which one’s subject is living at some significant cultural/historical distance from oneself, this matching phase may require a lot of the simulator. The pretend-beliefs

- at issue may be a pretty hefty set of beliefs of the sort one might find represented in a weighty anthropological monograph.^{viii}
- (2) The Simulation Phase, in which one's cognitive processes are put to work employing the content (the make-beliefs, make-desires, attitudes, and likely observations) assembled in the matching phase. This is the phase in which one determines where those contents would lead (assuming an agent were using processes relevantly similar to one's own). In contrast to an explanation in which one determines where the content would lead by ascertaining what a set of theoretically indicated transitions would yield, in simulation one merely puts one's processes to work on that information. One might then also label this stage more generically as *the one in which one determines where the content leads*. This alternative label highlights the functional parallel between the simulation phase and the application of theory.
- (3) An Attribution Phase, in which what transpired in the simulation phase serves as a basis for the attribution of lines of thought and thus dependencies in thought to one's subject.

Proponents of simulation as an account of human understanding and explanation of other humans differ over the details.

Alvin Goldman, understands the three stages as issuing in an analogical inference from ourselves as a model of the subject to an explanation of what transpired in our subject. Goldman envisions simulation as turning on a kind of introspective observation recording what transpires in the simulation phase and then involving an analogical inference supported by a similarity judgment. Goldman's understanding invites the involvement of propositional information in the matching and attribution phases. Notably, general theoretical information can condition one's confidence that one has managed to make oneself over into a reasonably good model of the agent—and thus condition one's analogical inference. None of this necessitates a role for propositional information—"theory"—in the simulation phase.

Stueber seems relieved that Goldman has not allowed propositional information a place there:

It is important to stress that simulation theorists are able to recognize the involvement of psychological knowledge in propositional form in our mindreading

activities as long as the propositional knowledge in question does not *infect* the simulation phase or make it causally unnecessary for reading other minds. If that would be the case then simulation could no longer be understood as a knowledge-poor strategy, and it certainly could no longer be understood as a strategy that is centrally important to our mindreading abilities.^{ix}

The point seems to be that, were theory or “knowledge in propositional form” to “infect the simulation phase” then simulation would not be much distinct from explanation turning on theory. I will soon argue that matters are more complex, that “infection” or “contamination,” while not the norm, is at least common—and that this does not make everything a matter of theory application. We are, I think, considerably more subtle epistemic agents than warring dichotomizing camps would like.^x

The simulation purist thus presents us with a dichotomous choice between alternative processes, as in Table 1, below:

[TABLE 1 HERE]

Here we are offered two disjoint kinds of epistemic processes for understanding and explaining others’ thoughts and actions. Simulation is understood as the default process. It is understood as applying our processes to whatever we manage to cobble together in the matching phase as pretend inputs. If this does not work, we need to have recourse to the costly and unwieldy alternative of information-rich modeling. When simulation works, theory or propositional information may inform what input one feeds into one’s cognitive processes, but these work unencumbered in their full default glory. Thus the simulation phase is uncontaminated.

3. Cognitive Diversity: or Why Purity May be Hard

Let us now focus on a difficult question concerning the role of propositional information in simulation: *To what extent can propositional information call for one to modify one’s own default cognitive processes as these are put to work in the simulation phrase.* Remarkably little attention has been paid to this question. Stueber and Henderson are among the few who discuss the matter.^{xi}

It is undeniable that there is real variation in the cognitive processes that people tend to employ; there seem to be real variation in risk aversion, both regarding actions and conclusions, real variation in default heuristics and their regulation by wider processes, in inductive safeguards, in reliance on analogies, in sustained levels of attention, in memory, and in many other characteristics of peoples' cognitive processes. Some of this variation may have roots in biological diversity within any population. Much of it may be rooted in differences in the training and experiences individuals undergo. Of course, some of these differences are associated with cultural differences. Because the need to accommodate variation in cognitive processes in simulation is central to the argument of this paper, it is fitting to pause here to take a crude inventory the kinds of cognitive variation that need somehow to be accommodated.

Intrapersonal—at a time:

Each of us, at any given time, has a repertoire of cognitive processes.^{xii} These include various heuristics, various cautions or sensibilities regarding the application of these different heuristics, various “high-end” processes that may be selectively deployed as cases call for it. Given my repertoire at a time, when presented with a run-of-the-mill matter, I may employ simple heuristics (I might judge that the probability of my having some minor condition is the probability of that condition given its best readily accessible symptom—the take-the-best heuristic). However, were the issue whether I have some feared disease for which I have a positive test result, I may reach for my statistics book, ask about base-rates in the general population, and stand ready to do a calculation using Bayes' Theorem.

Intrapersonal—across time:

One acquires some of one's cognitive tendencies over time, with training. As a result one's repertoire of cognitive processes is different now from what it was earlier. My own background, which includes graduate training in political science, makes me ask different questions in deciding what to make of political or social news than I would have asked prior to that training, and I think, different from what I would have asked were I not to have received this training. It is not just that I have different information, some of my cognitive habits were also modified. Undoubtedly, some of one's cognitive processes are the result of less formal training—of inarticulate lessons learned in the course of one's life. Some of these will be informal cultural lessons—learned perhaps by watching culturally respected agents.

Others will be learned in the course of one's own inferential practice—with its attendant frustrations or vindications of expectations one acquires sensitivities to pitfalls and a confidence in one's cognitive practice. Thus, not only does one have a repertoire of cognitive processes that one can call into play at a time, but one's repertoire itself changes over time.

Interpersonal Variation—across time.

Thinking about intrapersonal differences in cognitive tendencies across time, and in particular of those changes one expects with cultural training (formal or informal), leads to thinking about variation in cognitive repertoires across people and across cultures. In this connection, one might reflect on what one hopes really is accomplished by training in a science—one wants it to make some significant difference in the cognitive processes instanced in the scientific community. Kitcher provides a stylized model for thinking of scientific progress in which iterated generations of scientists are trained in a range of then contemporary best practices.^{xiii} As a result, at least in professional contexts, likely in some others, the initiates think in ways somewhat different from what they would have thought without such training. These initiates apply those trained sensibilities in a work phrase in which what counts as best practices may itself undergo some change. Along the way, practicing scientists train the next generation of scientists in what has emerged as the new best theoretical and experimental practices. (This sketch is intended to leave room for some variation in cognitive processes instilled across the scientific community at a time—variation across schools and labs, and individuals.) Over iterations of this cycle, trained cognitive processes and sensibilities may undergo significant change. This basic idea seems broadly applicable beyond the history of science. Training into a community involves (among other things) learning to think in certain ways—acquiring cognitive sensibilities and processes that are marginally different from what they would have been otherwise—and experience when thinking in these ways in a community sometimes results in revised sensibilities, and somewhat different training for later initiates.

Interpersonal Variation—at a time.

The cultural training just mentioned makes for differences in individual cognitive repertoires across time, and differences in community repertoires across time—and these differences across communities amount to differences across individuals in cognitive repertoires at a time. Again such differences need not be the result of formal training, as

cultural learning is quite diverse. Further, there may be differences in the cognitive tendencies of individuals that are not cultural in their origins. There may be differences in individual cognitive tendencies that have subtle roots in individual biological variation or early individual learning that is far too unsystematic to be thought of as cultural in any extended sense. Such differences are not eliminated in cultural training. Again, Kitcher suggests that differences across individuals make for a mix of cognitive strategies that fit within the cultural-scientific community standards and that such variation may be cumulatively beneficial.^{xiv} I would doubt that contemporary scientific practices are distinctive in this respect. There is significant variation both within communities and across communities.

Accommodating Variation in Simulation.

Central to my thinking is the following: *When one possesses information about such variation in cognitive processes, and when this information indicates that one's subject employs (or likely employs, or easily could employ) cognitive processes that differ somewhat from processes that one would normally employ oneself, this information should be accommodated in the course of one's simulation—or one should there abandon simulation.* This follows from the central supposition of simulation—on all accounts simulation relies on one's own cognitive processes (those employed off-line) being similar to the process in play in the subject simulated. Supposing that one is to use simulation in the face of marginally different cognitive tendencies on the part of oneself and one's subject, information about your different cognitive processes should be accommodated in the matching stage in a way that then conditions what transpires in the simulation phase. How this can be managed, the extent to which it can be managed, and the limits to accommodating cognitive diversity in simulation, will be the motivating concern through the remainder of this paper.

There are several broad ways that one might think to accommodate information about variation in cognitive processes within a broadly simulationist framework. Let us first consider how one might attempt to do so within the constraints of what I am calling simulation purism—again *simulation purism* is the claim that there can be no theory involvement in the simulation phase. The simulation purist insists that, if theory plays a role in this inner sanctum of simulation itself, the whole thing is “contaminated” or “infected” and one might as well not be a simulation theorist.^{xv}

Ultraorthodox Simulation Purism

On reflection, the simulation purist might imagine that information about cognitive diversity in processes can be accommodated in one of two ways. First, most restrictively, the simulation purist might suppose that within the bounds of what is properly termed “simulation,” one can only respond to such differences by taking on certain pretend-meta-beliefs concerning desirable cognitive processes. The idea would be that, where one’s subject has acquired a tendency to reason in such and such a way, different from one’s own cognitive tendencies, one might merely pretend-believe that such-and-such is a good way of reasoning. If one thinks that such moves in the matching phase is all that simulation can do by way of accommodating cognitive diversity, one is what I will term an *Ultraorthodox Simulation Purist*. On this understanding of simulation, if simulation is to accommodate cognitive diversity, it would need to do so by way of pretend-meta-beliefs, as in Table 2.

[TABLE 2 HERE]

Orthodox Simulation Purism

Less restrictively, a simulation purist might envision accommodating some cognitive diversity in a second way—a way that still largely turns on the matching phase. It might be allowed that, in the matching phase, one might prime oneself to put into play some select range of one’s own repertoire of cognitive processes—so that this set is relied on in the ensuing simulation stage. The purist might reassure himself that it remains true that one merely puts one’s processes to work on pretend input—only now it is just some subset one’s cognitive processes that are at work on the pretend-input. One might call this Orthodox Simulation Purism.

I first argue against Ultra-orthodoxy. In subsequent sections I will provide reasons for thinking that Orthodoxy is itself too restrictive. As I have stated, I believe that simulation is a very important capacity and practice in our epistemic toolbox—one that is commonly crucial to our ability to interpret and understand others. But, I have little sympathy with this simulation-purist bright line. I believe that the line is sometimes breached. It can and must be breached a lot in some cases, in which case no one would want to call the interpretive or explanatory practice there a case of simulation. But, in other cases, it may be breached just a little, in which case one should be free to talk of simulation. There is, I argue, a range of intermediate cases. Our interpretive and explanatory practice is flexible and

nuanced. We understandably have a preference for information-poor strategies (notably simulation), and this preference makes good epistemic sense. For critters like us, there is no real alternative to the common reliance on simulation. But we do can and do use mixed strategies, information-poor, information-moderate, information-richer, and information-rich, strategies as the case calls for them.

4. Against Ultraorthodox Simulation Purism

Cognitive diversity cannot be adequately accommodated in simulation by taking up different beliefs about how one ought to reason—it cannot be adequately accommodated just by pretend-meta-beliefs. Differences in cognitive processes are not simply the result of different beliefs directing one’s cognitive processes. Commonly, one’s cognitive processes are in significant measure the result of long training. In the course of this training, one can come to possess information and sensibility that is essentially richer than what one can dredge up in the form of paradigmatic beliefs. In this respect, one’s cognitive life, including one’s deliberation and reasoning processes, is not much different from one’s perceptual processes. Perceptual processes can serve as a simple illustration of how one’s cognitive processes can be shaped and colored by rich information, some of which is sub-doxastic (not encoded in paradigmatic beliefs). One’s perceptual processes began taking shape early, and all along the way, in uncounted episodes, one acquired rich information in connection with vindicated expectations and frustrated expectations. Thus, if you are like most of us, in the first 6 months of your life, you made real gains in your ability to perceive unitary but significantly occluded objects. Clearly, you were learning. But, it also should be clear that this did not largely consist of the acquisition of a set of meta-beliefs characterizing when you were to conclude that you had one irregular object, two more regular objects with one occluding the other, three objects touching but none occluding the others, and so on. And, of course, your perceptual processes continue to develop today—still in response to frustrated or vindicated expectations. So do mine. As a result, each of us differs significantly from the other in some perceptual domains. You are a lovely perceptual wonder, and so am I. With respect to either of us, it would be silly to think that one manages one’s perceptual chores by applying meta-beliefs expressing how it is best to move from input to a verdict. *Our informed processes are essentially richer than our meta-beliefs.* I believe that much the same can be said for one’s theoretical reasoning—for example. One was exposed to many

examples, to many exercises, to many vindications and corrections at the hands of teachers, and at the hand of events. We have each learned a lot more than either of us can say. I do not manage my cognitive chores by following a doxastic rule-book or by a prudential application of a set of meta-beliefs. Neither do you.^{xvi}

The core of the simulation story is that one is to get an explanatory sense for the dependencies obtaining in one's subject's thought by realizing in one's own processes, there at work on pretend-input, the parallel dependencies. If one's own processes (and their products) then turn on certain pretend-beliefs, one is supposedly entitled to say that the subject thought or acted in the relevant way because they had a belief with the same content as the pretend belief. If, however, one's simulation turned on certain pretend-beliefs, and one's subject had or used no such pretend-belief, something has gone wrong. This is no less true when one's pivotal pretend belief is an epistemic meta-belief. If one's subject did not have or use it, they did not think as they did because of it.

In effect, Stueber appreciates that ultra-orthodox simulation purism would be extremely limited and problematic, because it would need to accommodate cognitive diversity by deploying epistemic meta-beliefs. He writes:

[O]ne should agree with the social theorists like Pierre Bourdieu for whom normative social practices rest on an embodied but nonrepresentational sense of the “feel of the game,” which allows agents to act appropriately in their social world. Bourdieu uses the notion of a “habitus” to refer to an agent's embodied and practical know-how, which enables him or her to participate in social practices. The habitus is the “set of structured and structuring dispositions” that provides us with a scheme of perceiving the world and also with the appropriate emotional reactions to the world in its infinite variety. As Bourdieu stresses, the acquisition of such a habitus should not be seen as the outcome of explicit theoretical instruction but is primarily acquired by growing up in and becoming practically familiar with social practices through trial and error, imitation and mimicry of the competent players of the game.^{xvii}

Stueber here recognizes that deep, learned, implicit information—making for structuring dispositions—may vary across cultures. I would emphasize that something of this nature is significant in our epistemic lives, as it is in our other endeavors. Stueber is right to note that

such cultural variation poses significant challenges for one who seeks to rely on simulation in such contexts—and it seems that ultra-orthodox purist simulation would be particularly limited just here. However, in appealing to Bourdieu’s conception of *habitus*, Stueber seems inadvertently to understate the challenge faced. This is because, Bourdieu’s notion of *habitus* supposes that deep, implicit, structuring information is *shared within* a culture—the same information is supposedly shared across all the competent members of the relevant culture. There are reasons to doubt that this is so. If there is significant variation in the structuring information possessed by agents within a culture, then similar difficulties will attend simulation within a culture as well as across cultures, and the ultra-orthodox simulation purist will be particularly hobbled. Stephen Turner tellingly criticizes the idea that there is some one thing—one set of implicit structuring information—shared by all who (competently) participate in some cultural practice.^{xviii} Turner draws on connectionist cognitive science (as would I) and is alive to the importance of information that is possessed by trained agents, while being nowhere occurrently represented in their respective cognitive systems.^{xix} Differences in such information possessed across agents will always make simulation difficult, or “bumpy” at best. This is what Steuber correctly registers. However, what Turner argues is that, contra Bourdieu, there is significant variation in such information to be found across competent agents *within* a given culture. Turner pursues these issues in a particularly clear headed and revealing way.^{xx} Further, insofar as something along these lines holds true of us, and makes for aspects of human cognitive diversity, it would be inappropriate to try to accommodate that diversity simply through the contrivance of pretend epistemic meta-beliefs.

There is a paradoxical aspect to ultraorthodox simulation purism. Simulation theory has long been motivated by the realization that one’s understanding of how we humans reason is significantly less articulate than one’s ability to understand why various folk acted or thought as they did. In many respects, what one can dredge up as an articulate belief about how folk reason seems less powerful than one’s ability to predict what folk will do, or how they will think. Recognizing this paucity of supporting theory strongly motivates the simulationist insistence that one must commonly rely on one’s own processes (rather than theoretical generalizations) when understanding or explaining others. Now we have observed that there is significant cognitive diversity—significant diversity in the cognitive processes that an individual may have in play across times, diversity in the processes across individuals and cultures. What is paradoxical about ultra-orthodoxy is that it would try to

accommodate this diversity by supposing that it is paralleled by epistemic meta-beliefs. This amounts to the idea that the differences across individuals can be understood in terms of doxastic meta-representations—meta-representations, not of how folk reason, but of how one ought to reason. I will wager that one’s ability to articulate how one ought to reason, one’s ability to thematize such matters at the level of belief, is about as limited as one’s ability to characterize how people do reason. My doxastic take on how I ought to reason is, I think, about as impoverished as my doxastic understanding of how I—or others—do reason. My understanding of others’ doxastic take on how they ought to reason is obviously even worse off, and certainly worse off than my take on how they do reason. To suppose that one could accommodate cognitive diversity by finding and attributing a regulative set of meta-beliefs is tantamount to supposing that folk had articulate system of normative meta-beliefs just where the simulation theorist supposes that we all lack a (articulate or inarticulate) system of descriptive psychological beliefs.

The upshot: there must be ways of accommodating cognitive diversity in simulation that are not a matter of simply taking on pretend epistemic-meta-beliefs. In the matching phase, one must undertake *to simulate by matching processes in a way that runs beyond matching beliefs and desires with pretend-beliefs and pretend-desires*.^{xxi} How this works, and what it says about how propositional information—“theory”—can enter and condition simulation, can best be pursued by reflecting at length on the simulation phase itself and on how it may be conditioned by information accommodated in the matching phase in ways that are not merely a matter of taking on pretend meta-beliefs.

5. Cases, More Cases, and the Limits of Purist Simulation

Orthodoxy in focus

Having repudiated Ultra-orthodox Simulation Purism, we now explore the somewhat less restrictive position that I above called Orthodox Simulation Purism.

Each of us has a repertoire of cognitive processes. The exact extent of variation in the range of cognitive processes an individual employs across cases is an empirical matter—but it seems clear that one is not always on one’s best cognitive behavior, that one

sometimes employs a more sophisticated mixes of cognitive processes than at other occasions.^{xxiii} The simple point is that one can deploy less than one's own full repertoire of cognitive processes when deploying one's processes in a case.

Now, if some of one's cognitive repertoire is the result of training, and others lack that training, then one should not expect those others to exhibit the range of processes that one can exhibit. Similarly, if folk sometimes use only the more rudimentary components of their full cognitive repertoire, and if folk (including oneself) sometimes use less of their sophisticated repertoire than is on reflection "called for," then one should expect that others (even others with a cognitive repertoire at least equal to one's own) sometimes use less than their full cognitive repertoire, less than one's own full cognitive repertoire, or less than one finds "called for" on one's own reflection. This common kind of intrapersonal variation in folk use of their cognitive repertoire, would seem to call for a form of simulation in which one puts one's processes to work on pretend input *without thereby employing the full range of one's cognitive repertoire*. In some cases, it seems, one would best simulate others by deploying fewer processes than one would normally be inclined to deploy, given such input. This seems exactly what would be called for when simulating folk who have not acquired by training the full range of processes included in one's own cognitive repertoire. It also seems called for when one's subject is contextually prompted to apply some limited range of processes that correspond to some limited range of one's own processes.

Here is an example of such cases of simulation. Perhaps one has so trained oneself that when considering a set of observations/experiences, and being tempted to generalize from these to a population, one routinely is sensitive to possibilities of sample bias. Thus, one's composite processes of inductive generalization include a check for sample bias—and one's sensitivity to sample bias exhibits a range of trained sensitivities. Suppose that one has propositional information to the effect that certain folk are not so sensitive. Now suppose that one's subject is one of these folk, and likely does not then have the kind of sensitivity to sample bias that is a part of one's own cognitive repertoire. One's subject starts of some observed cases—perhaps a few salient cases. But, one's default way of thinking on make-beliefs representing these salient cases would not yield a generalization regarding the matter; the full standard array of one's own cognitive processes would not issue in a generalization given such input. Still, *ex hypothesi*, predictably, one's subject did generalize. To run the simulation, one needs to selectively disengage one's own bias

checking processes, or elements of it, and use the remainder of one's processes in the simulation. Of course, typically, to apply a restricted repertoire of one's processes, one would often need to suppress some of one's cognitive tendencies.

The Limits of Orthodox Simulation

As all simulation theorists should acknowledge, alternatives to simulation are needed insofar as there are limits to the thoughts and actions that can be explained simply using one's own cognitive processes. In particular, insofar as folk manage to deploy cognitive processes that are not within one's own cognitive repertoire, we cannot simulate them merely by adding pretend-meta-beliefs or by calling into play (perhaps some subset of) one's own cognitive processes. In his final chapter, of *Rediscovering Empathy* (entitled "The Limits of Empathy") Stueber discusses such cases. He recognizes several respects in which simulation is limited. First, at any time, it is limited by one's repertoire of cognitive processes, as I have just indicated. Still, if one has the time and energy, one might undergo some training so that one adds to one's repertoire of cognitive processes. Thus, in principle, one might say that, given enough time and energy, simulation is limited, not by one's present cognitive repertoire, but by the plasticity of one's repertoire. For example, when one's subject is a member of a significantly different cultural group, and where that group learns cognitive processes that are very different from those in one's own community, one might conceivably "go native." For Steuber, this represents the limits of simulation. The limits to the elaboration of one's cognitive repertoire may be unworkable in the case at hand—as having undergone one form of cognitive training, one may then lack sufficient cognitively plasticity to cultivate cognitive processes of the sort that one's subjects have acquired,^{xxiii} or one may simply not be afforded the time for the acquisition of such new habits of mind. If, for whatever reason, this self-transformation in the service of orthodox simulation is foreclosed to us—we must abandon simulation and turn in the case at hand to theory-based approaches. In so saying, Stueber clearly espouses what I am calling Orthodox Simulation Purism.

As Stueber also recognizes, even when one could transform one's cognitive repertoire in the service of simulation, there can be several reasons to hold back. These represent further limits on simulation.

First, once one's cognitive tendencies have been formed along certain lines, it may be very difficult or costly to undergo a form of training that would make one over into a thinker with a cognitive processes that might serve as a model of those one seeks to understand. Even where it is in principle possible, it might be prohibitively difficult and take years of sustained retaining. We might be afforded the time, but reasonably prefer to do others things with the time.

Second, even if one could retrain oneself (perhaps by cultural immersion) so that one acquired those processes that one's subject deployed but which one had previously lacked, it is not clear that one would always want to do so in the service of a simulation-based understanding of the other. There can be costs here aside from the time and energy involved. Plausibly, one might loose cognitive processes that one values in the bargain. Further, there may be cases in which one would not want for oneself to have acquired those "habits of mind." Suppose, for example, that there are kinds of phobia that characteristically arise from some early trauma (I really do not know how plausible it is that there are). Suppose that one's subject have such a phobia. Of course, it might be difficult to acquire these phobias—an instance of the first worry. But suppose that one could acquire these phobias. Would one really want to become agoraphobic, or claustrophobic, so that one could simulate one with these phobias central in their cognitive repertoire? Or would one want to foster homophobic or xenophobic processes in one's self in order to simulate others? I think not. Stueber apparently agrees:^{xxiv} we resist acquiring cognitive habits that are themselves epistemically, socially, or morally problematic.

Stueber discusses a related limitation on simulation: he discusses *a resistance to simulating by deploying cognitive processes that are within one's repertoire*, when those processes, as there applied, would be morally (or presumably socially or epistemically) problematic. It is easy enough to recognize the phenomena of "imaginative resistance" to which Stueber our attention.

However, I want to mention one curious element of the discussion. Stueber is concerned with Browning's treatment of German Reserve Police Battalion 101—drawing on Milgram's and Zimbardo's work.^{xxv} He thinks Browning's treatment is plausible, as do I. But, he is insistent that such treatments should be unambiguously categorized as a matter of the application of theory, and not as involving simulation. Perhaps this is correct. What is less clear is that this need be the case, or why it need be the case. On my reading, Milgram shows

us that certain tendencies to comply with authoritative requests or demands are in the cognitive repertoire of pretty much each of us—and that they can be very hard to resist.^{xxvi} Even those experimental subjects who resisted the instructions in Milgram’s experiments, did so with difficulty and stress. (Indeed, the capacity for such resistance seems less well developed in many of us.) As a result, one can wonder why it might be difficult to simulate Milgram’s subjects, or the members of Police Battalion 101 for that matter. Stueber seems to suggest that doing so is a matter of moral repugnance—and perhaps is destructive to our self-conception. I have two reactions.

We may find simulation of others in the contexts in question difficult for reasons that are more general, and perhaps less dramatic, than the imaginative resistance Stueber posits.

It may be very difficult, perhaps beyond one, to really take up the relevant inputs as effective pretenses. In one study, participants were provided a verbatim reenactment of an episode involving one of Milgram’s subjects. The reenactment terminated just before the subject actually administered the maximum shock to the “victim.” When then asked to predict how other subjects would behave, the participants in this study consistently and dramatically underestimated the degree to which subject of Milgram’s experiment would yield to the demands—although the situation of Milgram’s subjects had been made quite clear to these participants.^{xxvii} But, is this a case of an imaginative resistance to simulating Milgram’s subjects—is it a case in which we share processes but recoil at the prospect of indulging them in this repugnant fashion. There is reason to understand in other terms the failure of simulation seen here.

Goldman discusses the difficulty of simulating the “endowment phenomena” (see also Nichols and Stich [2003])^{xxviii}. The endowment effect is the robust tendency for people to rapidly come to set a significantly higher price for an object if they actually own it, than they would if they did not own it. After allowing subjects to examine a coffee mug Loewenstein and Adler asked one set of experimental subjects to predict what dollar amount they themselves would insist on in selling that mug *were* it theirs.^{xxix} They were later given the mug (as were a second group who had made no predictions). It turns out the average price predicted by subjects when they had been unendowed was significantly lower than the average prices subsequently set by either group of subjects once they came to be endowed.^{xxx} Thus, folk have difficulty simulating their future selves—but clearly this is not

because they lack the cognitive processes that are subsequently engaged (upon shortly coming to possess the item). Apparently either some “full-blooded” pretenses are not easy, or some processes in one’s cognitive repertoire are not easy to engage on pretense. When this obtains, prediction will be difficult or impossible, even when cognitive repertoires are relevantly similar. Goldman thinks that while it may be easy to imagine owning the item, it may be somewhat harder to really take up the pretend-contents fitting to really possessing the item. The relevant pretense is not merely a matter of imagining—he insists.

If this is correct, then it would seem all the more difficult to take up in pretense the situation of one in a destructive obedience study, or in a highly stressful real-life situation. It may be hard, not because of it is morally troubling, but because emotionally powerful simulation is hard generally.

If we can, simulating these things (from the minor to the monstrous), might be a good thing, and thus on balance should not be resisted. If these tendencies are in each of us (or most of us), there are reasons for undergoing the simulation. It might disabuse us of our convenient belief that we are really so different. Thus, it is said that Americans commonly hold that they are just not like those who torture, massacre, and the like. They commonly attribute things like Abu Ghraib and Mai Lai to “a few bad apples.” Perhaps some sustained simulation would convince them—or parallel others—that they are not so unique or exceptional. Understanding others—even others doing horrible things—can be good for us. Perhaps it can prepare us to better resist these tendencies in ourselves.

The truth is that I am not at all confident about these cases. If simulation of such emotionally fraught episodes is possible, and if it would disabuse folk of their exceptionalist tendencies, it would in that way a good thing. On the other hand, I can envision other possibilities. Bicchieri discusses how “rules” or “norms,” and expectations for conformity or violations can be primed in a context.^{xxxii} It is not at all clear to me whether simulating Milgram’s subjects or the thoughts and actions of various members of Battalion 101, would enhance one’s determination to resist tendencies that may lie deep in oneself, or instead would enhance the expectation for widespread norm defection. On Bicchieri’s account, the latter would undermine the authority of the norms against such horrible deeds. I welcome empirical data on this matter. Human psychology holds many surprises, and armchairs are no protection.

Returning to the main line of thought, this discussion of the limits one might

envison for simulation as understood by the Orthodox Purist, exemplified by Stueber, has shown that simulation in the face of cognitive diversity will be limited by (a) the plasticity of one's cognitive processes, (b) the difficulty of acquiring new cognitive processes even when one could, (c) the desirability of acquiring certain processes that are within others' repertoires—notably when those processes are themselves epistemically, socially, or morally problematic. It may also be limited by the difficulty of really taking up a rich enough pretend states to deal with certain kinds of emotionally or motivationally rich cases. In the face of these limitations, when one has reason to think that the processes in play in one's subjects are not among those in one's cognitive repertoire one will need to understand one's subjects in informationally rich—theory-driven—ways (at least until one has retained oneself—if that is desirable). Such is the Orthodox Simulation Purist view (as in Table 4):

[TABLE 3 HERE]

It is worth noting that this full Orthodox Purist view continues to draw a bright line between simulationist episodes and episodes that apply theory, and that what makes for the difference between the two remains that difference first represented in Table 1: the envisioned purity of the simulation stage itself. If one puts one's processes to work on pretend input, its simulation. If, instead, one figures out where some processes, propositionally characterized would lead an agent, given certain beliefs and desires (as initial conditions), its not simulation.

6. Grey Tones: Beyond Purism

I now want to take inspiration from two ideas.

The Pervasiveness Micro-Simulation in Common Theory Application

The first is a suggestion that I loosely take from Jane Heal. The idea is that there is always a little simulation going on even when one applies one's theory in what amounts to paradigmatic theory-theorist ways. The vast preponderance of theories of human psychology of the sort that plausibly are applied to explain actual thoughts and actions—at least those applicable in most social sciences, in history, and in everyday context—treat of the relations between contentful states. They characterize inferential moves involving

contents—they are not at the level of syntactical transformations. They may treat of “obvious inference” and of one’s concerning matters regarding which people are highly interested and attentive. They may say that people are inferentially better on “concrete familiar matters,” or posit facility with “cheater detection.”^{xxxii} They may say that people use various heuristics, in which they deploy a sense for “representativeness,” or key on what is “readily accessible.”^{xxxiii} They may deal with fast and frugal inferential strategies such as that using differential “recognition” (typically name recognition) as a crude indicator of relative size of cities, or of prospects for relative growth in stock value between companies).^{xxxiv} The theories may treat of “strong inclinations” to “comply with instructions of those recognized as authorities.”^{xxxv} One could go on here at great length generating examples (I will look a little closer at a few examples below). But, just now, I want to draw attention to this point: To apply such theories to the case of a situated agent, one needs to manage to follow out the contentful transitions indicated, beginning with the contents that the situated agent would likely have had. To do this, one does and must, in some elementary way, put one’s processes to work on the relevant contents. One must recognize what would seem “representative,” what would seem a “statistically good indicator,” and one must be able to follow out the inferences as one’s “theory” indicates. One must recognize what kinds of features might plausibly be correlated with recognition, what kinds of rule violation makes for “cheating,” what would be a “concrete” or “familiar” matter, what kind of matters attract and hold peoples’ attention, what seems “simple” or “complex,” what it is to recognize someone as an “authority,” and what range of features might mark another as an authority within a group, what acts would display or assert “domination” in context, and so on. The upshot, put simply, is that our application of such theory typically turns on a little use of one’s own rudimentary cognitive processes—namely one’s ability to work with content.^{xxxvi} We use this ability to trace out instances of semantic transitions to which our theory indicates our subjects are given, in order to ascertain what particular beliefs, desires, or actions the theory indicates would follow from others that serve as initial conditions in the case at hand. Thus, even in the application of our standard psychological theories there remains what are, in effect, moments of mirco-simulation. I do not understand this as a radical claim, and I do not propose to do anything radical with it.

Apparently Intermediate Cases and a Framework

My second inspiration arises from reflecting on cases in which it often seems very hard to say honestly whether one is simulating or applying theory—cases that seem to be intermediate between those at the limits of simulation (according to Orthodox Purism) and those that involve theory application in that crucial stage in which one determines where one's subjects' contentful input would lead them. For the simulation purist, there can be no intermediate cases.

Just on the simulation-theoretical side of the space to be explored are cases in which theory (or at least propositional information) about how people (or a group of people, or an individual) tend (tends) to reason informs one's subsequent simulation by indicating in the matching phase that one needs to prime oneself to employ in the simulation phase a select subset of one's own cognitive repertoire. Here, information about the cognitive repertoire of one's subject or subjects, and about the range of that repertoire that is likely to be in play in the episode of interest, informs the matching phase. Still, in the simulation phase itself, just one's own cognitive processes (or some subset of them) are in play. Thus, the simulation phase itself remains uncontaminated. *Here, the transition from input contents to output content (and action) is realized only by deploying (perhaps a select set of) one's own cognitive processes.*

On the theory-theoretical side of the space to be explored are cases in which theory or propositional information indicates that some of the agent's processes are not among those that one has in one's own cognitive repertoire. At these points one needs to apply theory—or a propositional characterization of the agent's processes—in light of a characterization of initial conditions. To apply the theory, one may need to employ some of one's own rudimentary cognitive processes to determine what the indicated cognitive transitions would yield in the case at issue (one may need to engage in theoretically directed micro-simulations), but (at least in the paradigmatic case) all the transitions are indicated and directed by theoretical understanding of the relevant processes. *Here, the transition from input contents to output content (and action) is realized only by determining what the propositionally characterized processes would yield.*

Given the above, one can say what would amount to an explanatory engagement that is intermediate between the simulation purist sort of approach and the largely theory-infused approach. Abstractly, the intermediate explanatory practice would be a matter of how one determines in the pivotal second stage what particular contentful output (decision or

conclusion) would be yielded, given supposed input. It would involve tracing out some theory-indicated transitions there, but would also turn on significant use of one's own cognitive processes to track transitions that are themselves at best crudely specified. Specifically, the transition from input-contents to output-content (and action) would be:

- (1) tracked in *some* measure by determining what the propositionally characterized processes would yield, and
- (2) tracked in *some* measure by deploying a select subset of one's own cognitive processes, where
 - a. one there tracks the transitions by employing cognitive processes that themselves go beyond rudimentary one's for working with contents, and thus
 - b. not all the use of one's processes are merely micro-simulations under theoretical direction.

What interests me are cases in which the applicable theory is "thin" at points, in a way that amounts to a significant underspecification of some of the cognitive processes in question—thus inviting or necessitating the reliance on stretches of simulation, while at the same time, the shared cognitive repertoire is sufficiently limited that one must instead rely on the application of theory to understand other moments in the same episode of cognitive processing. When this obtains, one must rely on both theory and simulation in a complementary way that does not fit comfortably within simulation purist approach or within view that appeal only to theory with micro-simulation in its application.

Here is one plausible example of such an intermediate case. Historians and others have drawn from Zimbardo's prison experiment lessons about the kinds of cognitive processes that one should expect to emerge among prison guards and other's with similar authority and power. These propositional/theoretical expectations have been used in attempts to explain what transpired in the case of German Police Battalion 101, for example.^{xxxvii} On the Stanford Prison Experiment web site, one reads:

There were three types of guards. First, there were tough but fair guards who followed prison rules. Second, there were "good guys" who did little favors for the prisoners and never punished them. And finally, about a third of the guards were hostile, arbitrary, and inventive in their forms of prisoner humiliation. These guards appeared to thoroughly enjoy the power they wielded, yet none of our preliminary

personality tests were able to predict this behavior.^{xxxviii}

Applying this to some episode seems clearly to call on one to use one's own cognitive processes (to simulate) in order to gauge what a guard of the happy-power-wielding type might think to do on a particular occasion. In doing so, one needs to be able to recognize what would count as instances of dominance behavior in their "inventive" variety from the point of view of the guards in concrete recorded situation. There are patterns to be found here—and characterizations of those might be used as somewhat more specific bits of theory. For example, one might note a tendency to humiliate one's charges, and even a connection between forced violations of sexual norms and humiliation.^{xxxix} But, even here, it seems to be that the lessons on offer remain "thin," so that applying them to concrete cases in their inventive "glory," turns on one's using more than rudimentary cognitive capacities. The theory seems to direct us to see the case in certain terms, but doing so seems to call for significant moments of simulation that are not aptly seen as mere theory-directed micro-simulation. The simulation called for not simply a matter of using rudimentary capacities for working with content to follow out theoretically well-specified transitions.

It is worth noting that only about a third of the guards were of the happy-power-wielding sort in Zimbardo's experiment. It may be that two thirds of people have dispositions that lead them to make other moves in such situations. Perhaps to project what one of the happy-power-wielders would do in a situation, folk with these other sets of composite dispositions need to make some of the moves in theoretically directed ways—while allowing nontrivial transitions yet to be managed with those cognitive processes that they yet share with the power-wielders. Their understanding would then depend on stretches of simulation beyond micro-simulation and instances of theory application, both deployed to understand what may be a very quick episode.

It might be wondered how do I propose to distinguish clearly between:

- (i) The simulation-theoretical cases in which theory operates in the matching phase to select a portion of one's cognitive repertoire that is then employed in pretense in the transition-tracking stage,
- (ii) The intermediate case in which one applies theory in the transition-tracking stage by using non-rudimentary cognitive processes that are indicated by theory, and

- (iii) The theory driven kind of case in which one relies on theory in the transition-tracking stage—typically theory that specifies transitions in terms of contentful transitions and thus requires one to use one’s rudimentary abilities to work with contents.

If this meant to call for the drawing of sharp lines, I want to resist. My contention is that there is a continuum of cases and there are no sharp lines to be drawn. Given the range of cognitive diversity surveyed earlier, there will be episodes in which the processes in play in the agent that we seek to understand will be a composite of shaped cognitive processes—some of which one shares, some of which one does not. Where one to share only very rudimentary cognitive processes, allowing one to track contentful transitions described in some theory, we have a case of type (iii), in which micro-simulation is employed in the application of theory. Were one to have largely the same composite processes in one’s repertoire, we can manage to engage in simulation of a sort that the simulation-purist can love—a case of type (i). But there are bound to be cases in which the processes in play in the agent are not merely some subset of processes that one can selective call into play—and where some of the components are among those in one’s cognitive repertoire. To track the transitions that are the result of component processes that are not shared, one will need to direct one’s transitions in light of propositional understanding of these component processes. At the same time, the shared processes may afford moments of simulation that are not aptly understood as mere micro-simulation. Indeed, if one’s theoretical understanding of these shared processes is very thin, one may need to rely on simulation here. Confronted with such cases, one would typically find oneself tacking back and forth between imaginative pretense and general expectations—or theory. It seems disingenuous or flat-footed to see this as either simply simulation or wholly theory application. In the end, it is this tacking-back-and-forth between theory and pretense that I want to honor.

Consider a case in which one has information to the effect that:

- (a) one’s subject or subjects used (or likely used) a process that is composed of simpler processes,
- (b) one has at least some of the simpler processes in one’s cognitive repertoire, but one does not have the composite process in one’s standing repertoire.

Here is a plausible example.

It seems that folk commonly are strongly influenced in their causal thinking by judgments regarding what antecedent or environmental phenomena resemble the effect. They, “look for causes whose principle features match those of the effect.”^{xl} The phenomena, much remarked on by Mill, may be understood as turning in an overuse of the representative heuristic which, “allows the individual to reduce many inferential tasks to what are essentially simple similarity judgments.”^{xli} Perhaps, from long education, one’s tendency to rely on similarity judgments as a basis for uncovering general causal connections (for arriving at causal generalizations) has been significantly extinguished. One now insists that one’s causal generalizations be backed by statistical information and careful analysis (either by oneself or by one’s ultimate community informants). (Or perhaps one’s tendency to attend to similarity has been transformed into a heuristic for the discovery of causal hypotheses that are not believed until subjected to test.) In any case, it no longer seems natural or fitting to move from judgments about similarity to causal attribution. At the same time, one may still share many of the same tendencies to judge similarities. This adds up to a situation in which one’s subject may instance a composite process [processes yielding similarity judgments + process of causal inference on the basis of similarity] that one lacks—although one has the dispositions to those processes that give rise the relevant similarity judgments.

Now suppose that one’s subject is “feeling a little wimpy in certain delicate contexts involving certain very intimate moments”—and seeks a remedy. To put it less discreetly, the desired effect is sexual virility. The subject notices certain critters that have attributes similar to what he would like to acquire. Perhaps he notices the duration and expressive character of giant tortoise mating—“that is kind of what I am after,” he may think to himself. Similarity judgment in hand, he then infers that ingesting select tortoise remains will give him the boost he desires. Admittedly, that is not the move you would have made on the basis of his evidence—even if you do find tortoise liaisons to be similar to what one would find congenial (with respect to their duration, at least). You can appreciate the similarity between what the agent seeks and tortoise sexual stamina, and this reflects a shared capacity for a certain range of similarity judgments. But, one’s own repertoire of processes for causal reasoning does not include the representative heuristic as a way of getting to general causal beliefs. One can, of course, characterize the process, and one can use one’s own ability to work with contents to follow the inferential moves indicated.

In this case, some of what one does is a matter of putting one's own processes to work on imagined input, resulting in similarity judgments that parallel those made by one's subject. Yet some of what one does is a matter of applying a descriptive understanding of how one's subject tends to reason—tracing out the inferential transitions that certain processes would bring about, given the similarity judgments one imagines on behalf of one's subject. In making use of information about one's subject—in determining where the information leads—one employs a mix of simulation and theory. In applying theory, one employs one's capacity to work with contents in recognizing instances of theoretically indicated transitions.

Here is another plausible case. In the course of social change, the status, life prospects, and economic well being of specific classes of people may be adversely impacted. Their valued models of how one ought to live may cease to be widely esteemed. Their skills and products may be in less demand. They may sense that many things are changing in parallel ways, and not for the better. Social and economic change understandably produces discontent. Again, these processes are widely shared—and it is not hard to simulate the kinds of disease and discontent that are occasioned. Think of this as turning on a shared component within most people's cognitive repertoire. Now, it is possible to take certain changes as symbolizing the whole constellation of changes that one finds troubling or threatening. Thus, sipping lattes while reading the New York Times may symbolize much that is wrong with America to the short-order cook—or to others engaged in many productive occupations. Again, I suspect we all retain processes that allow us to simulate this much—how certain salient episodes or groups of people can come to symbolize wider social phenomena. But, apparently, some pursue this construction of symbols of discontent by “endowing” some convenient minority with all that is thought wrong. Further, they may come to believe that this minority—possessing in a distinct way these resented properties—are *the* causal roots of the hurtful social change. Arguably, there is a long European tradition of both “endowing” the Jews with a constellation of horrible traits and thinking that these folk were the cause of just what troubled them—and doing so particularly in times of significant social change. Arguably, in the United States, some have developed views of immigrants from Central and South America that are laden with parallel symbolic and emotional baggage—attributing to them just what troubles their America, and seeing them as the source of what troubles their America.

When understanding American or German xenophobes in the way just suggested, what are we doing? Clearly, we take as our departure from some observations of the conditions in which our subjects find themselves. We begin with our subjects' earlier ethnic, national, and professional self-images, their reduced situation, and the uncertainty they then confront. On this basis we may have some success simulating the threats they may then experience. But I suggest that this success in simulation is already managed by framing one's pretense by (thin) theory of a sort. We may also simulate how they might "roll" certain somewhat foreign and newly salient motifs, associated with the perceived winners in the social changes, into a symbol for the threatening or harmful changes. Few, I think, lack this faculty for symbol-construction. Most can simulate this much, using their native cognitive faculties—at least with some priming in the matching phase, and subject to the empirical constraint of needing to account for what threats one's subjects ultimately "see." But, what of the causal inference? What of one's subjects' inference to the effect that these Jews or immigrants *are the cause* of the preponderance of changes that so trouble them?

Here, I think, most will not immediately or without theoretical prompting make such inferences or transitions. Perhaps we do not do so because we find it hard to engage in the pretense in a full-blooded enough way (see earlier remarks on simulating Milgram's subjects). Perhaps it is because we have by training largely extinguished the relevant processes—or put in place inhibiting processes that are not readily suspended. In either case one would need to appreciate these inferences by applying theory, following the inferential moves that "theory" indicates. Of course, we need to use our capacity to work with content to see an instance of the connection or transition that is suggested by our background theory. Theory thus directs us to explore such transitions, and I doubt that this exploration involves only rudimentary capacities to work with content.

The cases just suggested are quite tentative and somewhat conjectural; so I do not want to put too much weight on them. While I find them plausible as partial illustrations of what I envision, what I intend is that they serve roughly to illustrate the mixed strategy that I am sure each of us employs, and sometimes must employ. It is a mix in which simulation as commonly conceived takes one only so far—where it dovetails with theoretical descriptions which are applied to fill in some of the inferential moves that are exhibited in the case. What I suspect is that in many cases investigators find themselves using such mixed strategies. If the case calls for a set of inferential moves—and most do—and if some but not all of these

inferential moves turn on processes that one has in one's cognitive repertoire, then one probably employs a mixed strategy. What is more, in doing this, one probably switches from theory to simulation and from simulation to theory automatically, as effortlessly as a sophisticated automobile will shift the power supplied to various wheels in slippery conditions. Or, better, the picture that emerges in many cases will be rather like a pointillist painting—with many dots of inferential color supplied by the investigator as simulationist and many by the investigator as theorist.

As a philosopher, I need a name for my position: STP or Simulation and Theory Pointillism. (I might also call it Interpretive Posi-traction, barrowing the name for the early automotive technology for automatically distributing power to wheels as needed).

7. Reasons and Rationality?

Karsten Stueber has argued that simulation must be central to our interpretive and explanatory practice because only it enables one to grasp the reasons of agents—and thus to understand agents—as reasons (good or bad). It alone facilitates appreciating their reasons from the engaged perspective, rather than from the detached, third-person, perspective of theorizing. I doubt that things can be as clear as this suggests.

As noted above, when applying theory in the understanding of an episode, one commonly engages in micro-simulations—at least this is so when the theory or informing generalizations is couched in terms of semantic transitions or intentional terms (such as maximization, satisfaction, compliance, or dominance). In such cases, one takes up the apparent inputs of one's subjects and follows out the transitions to which one's propositional information indicates they are disposed. In some cases one will think that the contentful bases on which one's subjects depended constituted inadequate reasons or flatly bad reasons. Presumably one would so think in the case of the search for tortoise-based aphrodisiacs. But it is not clear why one would have difficulty seeing this as a matter of reasons. Of course, it might be that in other cases, one will say that one's subject did better than one would have done yourself. One's subject may have seen in a moment the possibility for an effective chess move, while one would not have seen it for oneself. Seeing this possibility, effortlessly as it were, would presumably turn on a trained-up expert sensibility that one lacks. One may model the results using much more plodding processes

that represent crudely the constraints to which one's subject is sensitive. Here, one is clearly doing something more like computer modeling of a system than simulating by using similar processes. That said, one yet can appreciate the reasons to which one's subject is responsive, and recognize them as reasons, indeed very good reasons. Thus, insofar as the propositional information about how various folk reason (one's "theory") is stated in terms of contentful relations and transitions (the common format) then something like our own capability for recognizing instances of such transitions is called into play in the application of that theory to cases. If this is not simulation, or micro-simulation, it is yet enough to allow one to appreciate reasons, good or bad. Theory application is not antithetical to reason's talk.

When we consider the various examples above of cases intermediate to paradigmatic simulation and paradigmatic theory application, the same point is emphasized. When simulation cannot get one from inputs to outputs because one repertoire of processes does not include processes that are sufficiently similar to those at play in one's subject, one still seems able to understand one's subjects as having reasons different from what would have moved oneself. Thus, we can appreciate the reasons moving the afficionado of tortoise gonads. We can appreciate the reasons moving the chess master—although perhaps too late to do us any competitive good. We can identify with and understand the reasons that losers in social change come to fear and hate certain folk who symbolize for them the hurtful change they perceive. When simulation cannot get one from inputs to outputs because, while one can imagine the inputs, one cannot take them up as sufficiently vivacious pretenses to trigger what are similar processes in our repertoire, one can turn to theory. One apparently needs to do this in the case of the endowment effect, and arguably in the case of Milgram's subjects and some members of German Reserve Police battalions. But, so far as I can see, none of this robs them of reasons (admittedly inadequate reasons we think) nor of agency. Perhaps we can even recall episodes in which we ourselves behaved very badly towards someone, where we felt moved by the authority of a group of which we were a part. This is not simulating our subjects, because it does not get us from what we imagine is their situation to the horrible results to which they were led. But, it is at least recognizing very loosely similar processes within our cognitive repertoire. It is some personal evidence supporting a theoretical perspective on the episodes. This may not be enough of a first-person perspective involving the subject's own situation for Stueber, but to my mind it seems enough for understanding their agency.

TABLES

ST-TT alternatives	P1—Matching Inputs—Assembling a Provisional set of Pretenses or Initial Conditions	P2—Determining where the information leads—making transitions or tracking the indicated transitions	P3—Attribution / Explanation
ST-ish	Propositional information regarding others' different situation—informs the beliefs, desires, and related states taken up in pretense	Simulation Stage —One runs one's own processes, as is, off-line, on the assembled input.	If one concludes/decides to do as, plausibly, the agent did, then one sees the agent as doing so for the reason's operative in one's off-line cognition.
TT-ish	Propositional information (of others different situation) informs one's take on belief/desire as initial conditions	Propositional descriptions of cognitive processes (indicating types of cognitive transitions) are applied to the initial conditions.	If the T-indicated transitions would lead to conclusions or decisions that, plausibly, paralleled the agent's, then one attributes such contents a role in the agent's thought.

Table 1: Simulation Purism, generic

ST-TT alternatives	P1—Matching / Initial Conditions	P2—Determining where the information leads	P3—Attribution / Explanation
ST-ish	Propositional information—of others different situation—informs make-belief, make-desire, and related concerns: pretend-sense-of-urgency (for example.) Accommodate cognitive diversity by taking on normative epistemic meta-beliefs (as pretend-beliefs).	Simulation Stage —One runs one's own processes off-line, as is. The pretend-meta beliefs contrived in the matching phase inform episodic processes in the simulation phase—as one's base cognitive processes (similar across agents) operate on this pretend input.	If one concludes/decides to do as, plausibly, the agent did, then one attributes the agent as doing so for the reason's operative in one's off-line cognition.
TT-ish	Propositional information—of others different situation—informs one's take on belief/desire as initial conditions	Propositional descriptions of cognitive processes—indicating types of cognitive transitions—are applied to the initial conditions as indicated.	If the T-indicated transitions would lead to conclusions or decisions that, plausibly, paralleled the agent's, then one attributes such contents a role in the agent's thought.

Table 2 Ultra-orthodox Simulation Purism

ST-TT	P1--Matching	P2—Determining where the information leads	P3--Attribution
Paradigm S-ish	Propositional or observational information—of others different situation— informs make-belief, make-desire, and related concerns: e.g., make-sense-of-urgency	Simulation —One runs one’s own processes off-line, as is, using the make-beliefs, make-desires, make-senses.	If one concludes/decides to do what, plausibly, the agent did, then one attributes the agent as doing so for the reason’s operative in one’s off-line cognition. Of course, one’s simulation may not parallel what the agent seems to do. Then one looks for better make-... Overall: IBE
S-ish Accommodate cognitive diversity in two ways:	1. By taking on (as pretend-beliefs) normative epistemic meta-beliefs	The pretend-meta beliefs contrived in the matching phase inform episodic processes in the simulation phase—as ones base cognitive processes (similar across agents) operate on this pretend input.	
Simulation Deferred	2. By selectively inhibiting some of one cognitive repertoire so as to subsequently put in play processes that more plausibly mirror those in one’s subject	Propositionally informed use in S of selective ranges from one’s own repertoire of cognitive processes--	
	One defers the above until one manages to acquire some cognitive process not now in one repertoire	After acquisition, one can selectively deploy one’s new cognitive processes.	
	The use of processes within one’s repertoire may be forestalled by either limits in one’s ability to take up pretense with sufficient vivacity or by various forms of resistance to doing so. If not overcome, one must turn to T-ish approaches		
T-ish	Further possible consequences of information about cognitive diversity	T-indicates that <i>some subset</i> of the agent’s processes are not among those that one has in one’s own cognitive repertoire. At just these points one needs to apply theory as below.	As below
Paradigm T-ish	Propositional or observational information—of others’ different situation— informs one’s take on belief/desire as initial conditions	T-application Propositional descriptions of cognitive processes are employed given certain T-indicated types of cognitive transitions.	If the T-indicated transitions would lead to conclusions/decisions that, plausibly, paralleled the agent’s, then one attributes such contents a role in the agent’s thought.

Table 3: The Full Accounting of Orthodox Simulation Purity

ⁱ I am indebted to Karsten Stueber, Jean Cahan, and Joe Mendola for discussions and criticism of ideas developed here.

ⁱⁱ A. Goldman, *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* (Oxford: Oxford University Press, 2006), chapter 7, provides a survey of the issues involved in both informed pretend-inputs and informed quarantining of beliefs and desires.

ⁱⁱⁱ K. Stueber, *Rediscovering Empathy* (Cambridge, MA: M.I.T. Press, 2006).

^{iv} There is reason to think that there is always, or at least almost always, a little simulation going on in the application of theory even in those cases in which simulation must give way to theory. The reasons are reflected in D. Henderson and T. Horgan, "Simulation and Epistemic Competence" in H. Kogler and K. Stueber (eds.), *Empathy and Agency: The Problem of Understanding in the Social Sciences* (Boulder: Westview Press, 2000), pp. 119-143 (as well as Stueber, *Rediscovering Empathy*). First, much of the propositional understanding that one might bring to bear on cases describes transitions between contentful states. Applying such understandings to cases requires one use one's own inferential capacity to make the transitions in contents indicated. To take a particularly simple case, suppose one's propositional understanding of how folk reason (absent special training) indicates that they are likely to use some heuristic—crudely gauge the probability of some event by whether it would be representative of the process understood to be in play, for example. To apply this to the case, one needs some take on how the process is understood by the agent, and some sense for "what results are like that process." One might unpack these things a little further, but, sooner or later, in applying this story one needs to use one's own cognitive processes to make the indicated transitions for oneself. Thus, for the application of much of the propositional understandings one is likely to find useful, one eventually gets to something like simulating. Second, many cognitive tasks with which one's subject might be involved require an agent to come to terms with an array of information that is relevant, where the agent must somehow appreciate what information is relevant. Understanding how this is done has come to be recognized as a deep problem. This is the core of the notorious frame-problem for classical, computational, cognitive science. We do not have a theory that describes how folk manage to put a frame around what range of information is relevant to various of their theoretical and practical issues—although we may have some crude generalizations. It then seems unavoidable to at some point rely on one's own processes—insofar as these are like the subject's—to get a sense for what information would strike them as relevant to the issue faced. I believe that the frame-problem is a deep, in-principle, problem that will frustrate purely computational approaches. As a result, we are unlikely to ever have theory that ultimately takes us beyond the need to simulate at those junctures where our subject must have solved the frame-problem.

^v Goldman, *Simulating Minds*, pp. 44-5, discusses this generate-and-test character of simulation, which he notes contributes to the hybrid character of simulation.

^{vi} Stueber, *Rediscovering Empathy*, p. 111.

^{vii} Stueber, *Rediscovering Empathy*, p. 111.

^{viii} Stueber, *Rediscovering Empathy*, p. 207; D. Henderson, "Conceptual Schemes After Davidson" in G. Preyer, F. Siebelt, and A. Ulfig (eds.), *Language and Philosophy: On Donald Davidson's Philosophy*. (Dordrecht: Kluwer, 1994), p. 171-97.

^{ix} Stueber, *Rediscovering Empathy*, p. 121.

^x Other writer's, notably Gordon, would have us think of the three stages differently. See, R. Gordon, "Folk Psychology as Simulation," in M. Davies and T. Stone (eds.), *Folk Psychology* (Oxford: Blackwell, 1995), pp. 60-73; R. Gordon, "The Simulation Theory: Objections and Misconceptions," in M. Davies and T. Stone (eds.), *Folk Psychology*. (Oxford: Blackwell, 1995), pp. 100-122; R. Gordon, "Simulation Without Introspection from Me to You," in M. Davies and T. Stone (eds.), *Mental Simulation*. (Oxford: Blackwell, 1995), pp. 53-67; R. Gordon, "Radical Simulation," in P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind*. (Cambridge: Cambridge University Press, 1996), pp. 11-21. Gordon sees the three stages as parallel to an actor's performance of a role. In the matching stage, one takes on a role in which one's character or subject has certain imagined beliefs. As one then simulates, one is already, in effect, attributing. Gordon thus insists that there is no need to make an analogical inference from what was observed of my own pretend play to my subject. Upon completion of simulation, upon adopting or resuming my own perspective, one has already, noninferentially, attributed the relevant states to one's subject. Gordon's proposal has the effect of minimizing the role of theoretical knowledge in the attribution phase. Analogical inference turns on judgments regarding the similarity between the agent's cognitive processes and one's own and it seems incontestable that much psychological information of a propositional character can and should condition such similarity judgments and one's confidence in one's simulation. Gordon seeks to minimize the place for such judgments, and thus such information, in the attribution phase. However, Gordon can ultimately get little mileage from these kinds of moves—issues of cognitive similarity must be faced squarely at some point. If one minimizes them at the attribution phase, insisting that they have already been addressed by what transpired in the unchallenged simulation, this can only be because somehow one supposes that there were addressed in thinking one had done a good job at the matching phase. The radical simulation theorist, such as Gordon, will end up playing "wack-a-mole" with similarity concerns—minimizing them at the attribution phase, but only to find them more urgent at the matching phase. I doubt that there is definitive answer to the question of just at which stage these issues are addressed—it may be a "pay me now or pay me later" situation. Still, in some cognitive fashion, they must be addressed along the way. Goldman's and Stueber's approaches have the virtue of not obscuring their importance.

^{xi} D. Henderson, "Simulation Theory vs. Simulation Theory: A Difference Without a Difference in Explanation," *Southern Journal of Philosophy*, Spindel Conference Supplement 34 (1996), 65-94.

^{xii} Throughout this paper, I am allowing myself to write intuitively and without great precision about cognitive processes being the same or being different. The basic idea might be clarified in various ways—some of this is the work of cognitive science—but I anticipate that my general lines of thought here would make sense on differing clarifications.

^{xiii} P. Kitcher, *The Advancement of Science* (Oxford: Oxford University Press, 1993).

^{xiv} Kitcher, *The Advancement of Science*.

^{xv} This language of "infection" or "contamination" is Stueber's, and he seems to have confronted most directly the issues posed by cognitive diversity (Stueber, *Rediscovering Empathy*, chapter 6). One can find some suggestive passages in Goldman's writings. Goldman, has written with empirical sophistication about the challenges posed by different belief-desire-sets in simulation. For example, he devotes significant attention to empirical

studies indicating a range of problematic projection effects among adults as well as children (Goldman, *Simulating Minds*, pp. 164-73), and to endowment effects and the difficulties of predictive simulation (Goldman, *Simulating Minds*, pp. 174-5). This leads him to distinguish the pretense needed for simulation from mere imagining. However, he has largely been silent on the challenges posed by variation in cognitive processes. At a few points, he does allude to the matter, as when he suggests in passing that in the set-up stage one might need to select not just pretend input, but also to “choose” a “suitable psychological mechanism” to deploy in the simulation stage (Goldman, *Simulating Minds*, p. 44). However, this suggestion comes when he is mentioning a form of hybrid simulation theory for which he expresses skepticism. All things considered, one can perhaps say this much. First, Goldman has not systematically addressed the need to accommodate in simulation the cognitive diversity in processes within and across human populations. Second, his openness to hybrid accounts would seem to allow for various augmentations of simulation along the lines that I soon explore. Third, at some points in his writings there is the suggestion that accommodating cognitive diversity in a largely simulationist mechanism would turn on one’s flexible and selective use of one’s own cognitive repertoire.

xvi For a discussion of how a cognitive system can, indeed must, employ information richer than the beliefs on which it operates—and for how such information shapes and conditions processes, see Henderson and Horgan, *The Epistemological Spectrum* (Oxford: Oxford University Press, 2011), particularly chapters 7-8.

xvii Stueber, *Rediscovering Empathy*, 208-9.

xviii S. Turner, *The Social Theory of Practices* (Chicago: University of Chicago Press, 1994).

xix S. Turner, *Brains/Practices/Representations* Chicago: University of Chicago Press, 2002).

xx S. Turner, “Mirror Neurons and Practices: A Response to Lizardo,” *Journal for the Theory of Social Behaviour* 37 (2007), 251-71.

xxi The point is acknowledged as a delicate and difficult one for simulation theorists by Stueber (*Rediscovering Empathy*, pp. 205, 208). See also Henderson and Horgan, “Simulation and Epistemic Competence.”

xxii Some of the empirical questions are reflected in the literature concerning whether or not it is proper to see folk’s cognitive processes as falling with two cognitive systems—one a system of simple heuristic processes (and related fast, cheap, and automatic processes), the other a system of more articulate (slower, more cognitively costly, deliberate) processes that may discipline the first system or revise its verdicts. See D. Kahneman and S. Frederick, “A Model of Heuristic Judgment,” in K. Holyoak and R. Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning* (Cambridge: Cambridge University Press, 2005), pp. 267-93; S. Sloman, “The Empirical Case for Two Systems of Reasoning,” *Psychological Bulletin* 119 (1996): 3-22; M. Osman, “An Evaluation of Dual-Process Theories of Reasoning,” *Psychonomic Bulletin and Review* 11 (2004): 988-1010.

xxiii It may be that no training would suffice. See Stueber (*Rediscovering Empathy*, pp. 210-11) recognizes these points, citing Henderson and Horgan, “Simulation and Epistemic Competence.

xxiv Stueber, *Rediscovering Empathy*, pp. 213.

xxv C. Browning, *Ordinary Men: Reserve Police Battalion 101 and the Final Solution in Poland* (New York: Harper Collins, 1992).

-
- xxvi S. Milgram, "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology*. 67 (1963), 371-378
- xxvii G. Bierbrauer, G. "Effects of Set, Perspective, and Temporal Factors in Attribution," unpublished doctoral dissertation, Stanford University, 1973.
- xxviii Goldman, *Simulating Mind*; see also S. Nichols, and S. Stich, *Mindreading: An Integrated Account of Pretense, Self-Awareness, and Understanding Other Minds*. (Oxford: Clarendon Press, 2003).
- xxix G. Loewenstein and D. Adler, "A bias in the prediction of tastes." *Economic Journal: The Quarterly Journal of the Royal Economics Society*, 105 (1995), 929-37
- xxx See also R. Thaler, "Towards a positive theory of consumer choice," *Journal of Economic Behavior and Organization*, 1 (1980), 39-60 and D. Kahneman, J. Knetsch, and R. Thaler, "Experimental tests of the endowment effect and the Coase theorem," *Journal of Political Economy* 98 (1990), 1325-4.
- xxxi C. Bicchieri, *The Grammar of Society* (New York: Cambridge University Press, 2006).
- xxxii L. Cosmides and J. Tooby, "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty." *Cognition* 58 (1996), 1-73
- xxxiii See for example R. Nisbett and L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall (1980). Another example: Various descriptions of heuristics appeal to the subject's sense for how representative a putative result is of the results that might derive from a given generating processes (for of a perceived similarity of causes and effects). It is said that one might explain an instance of the gambler's fallacy (in which one's subject bet heavily on the supposition that a result of some type was "due") in terms of thus heuristic. My point is simply that, to apply this propositional characterization of the inference tendency in this fashion, one must oneself be able to recognize what would seem representative—and one does this by using one's own cognitive processes to the case at hand
- xxxiv See, G. Gigerenzer, *Gut Feeling: The Intelligence of the Unconscious*. (London: Viking Publishing, 2007), G. Gigerenzer, *Rationality for Mortals*. (Oxford: Oxford University Press, 2008).
- xxxv Milgram's formulations, from, "Behavioral Study of Disobedience").
- xxxvi There are extreme cases in which one may not do this. As an example of the clearly theory-driven extreme, where one does not used one's own processes to track the indicated transitions, consider Oliver Sack's characterization of people who are prodigious calculators (O. Sacks, *Anthropologist on Mars: Seven Paradoxical Tales*. London: Pacador, 1995), pp. 182-3). Such people commonly (but not always) have a form of autism. As Sack's ultimately suggests (pp. 214-5) the processes by which these prodigies manage their mathematical chores are rather different from those most folk employ; they are not just quicker, for example. Notably, they also have a very different developmental profile. Were I called on to explain why such an agent arrived at a result, I might simple deploy the crude theory to the effect that the individual is good at the relevant transitions—and then plug in problem posed into a calculator.
- xxxvii Browning, *Ordinary Men*, pp. 167-8.

^{xxxviii} <http://www.prisonexp.org/psychology/33>, see also P. Zimbardo, *The Lucifer Effect: Understanding How Good People Turn Bad*. New York: Random House Trade Paperbacks, 2008.

^{xxxix} Zimbardo, *The Lucifer Effect*, pp. 171-73.

^{xl} Nisbett and Ross, *Human Inference*, p. 115.

^{xli} Nisbett and Ross, *Human Inference*, p. 7.