Simulating Metaethics:
Consensus and the Independence of Moral Beliefs

Daniel Simpsonbeck and Justin Sytsma[1]

Abstract: In a series of recent articles, Shaun Nichols (2019a, 2019b; Ayars & Nichols 2019) has called on statistical learning to offer a process vindication argument for lay metaethical beliefs. His argument proposes that people are sensitive to consensus information in forming metaethical beliefs and contends that calling on consensus information in this way is rational. This vindication of lay metaethical beliefs hinges on a number of substantive assumptions, however, including that people's first-order beliefs are to a large extent independent. In this paper, we raise concerns about this assumption. We argue that if people do call on consensus information in forming second-order beliefs, and if they maintain consistency between their first-order beliefs and their second-order beliefs, then independence will be threatened by the use of the very process at issue. This is tested through a series of computer simulations.

[1] We want to thank Richard Joyce and Shaun Nichols for their extremely helpful feedback on a previous version of this paper. We would also like to thank the audiences at VUW.

Some philosophers have held that lay people are moral universalists and reject moral relativism (e.g., Joyce 2001, ch. 4; 2011).[2] Some have attempted to explain away lay moral universalism, offering debunking arguments for lay belief, including process debunking arguments—arguments that belief in moral universalism issues from an epistemically defective process (e.g., Cameron, Payne, & Doris 2013). In contrast, Shaun Nichols (2019a, 2019b; Ayars & Nichols 2019) has called on statistical learning to offer a process *vindication* argument for lay metaethical beliefs. He argues that metaethical belief "is partly based on rationally appropriate processes, which points towards a possible vindication for the belief" (2019a, 99).

Nichols notes that recent findings suggest that there is variation in the extent to which lay people endorse universalism with regard to different claims (Goodwin & Darley 2008; Wright, Grandjean, & McWhite 2013). Further, this variation in second-order beliefs correlates with perceived consensus in first-order beliefs about the claim, with some evidence suggesting that perceived consensus plays a role in bringing about second-order beliefs (Goodwin & Darley 2012; Ayars & Nichols 2019). Nichols then argues that the process of calling on consensus information to form second-order beliefs is rational and, as such, that second-order beliefs formed in this way are (*prima facie*) rational. The idea is that in hypothesis selection, we should prefer hypotheses that more specifically fit the type of data that obtained, penalizing hypotheses that are overly flexible (i.e., that are able to fit a broad range of data that might have obtained). Nichols then contends that the hypothesis that the truth of a given moral claim is relative is able to fit whatever consensus data obtains, while the hypothesis that the truth of the claim is

[2] There is inconsistency in the use of metaethical terms in the literature, including within the papers we focus on here. Ayars and Nichols (2019, fn3) note this fact and write that "we use 'moral universalism' for the view that there is a single true morality, in contrast with 'moral relativism' which denies this." We will follow their usage, although taking the "moral" to be implied. Further, given variation in metaethical beliefs about different moral claims, we will apply the terms to beliefs about individual claims. Thus, we will use "relativism" to indicate the second-order belief that a given moral claim is true for some people and false for others and "universalism" to indicate either the second-order belief that a given moral claim is true for everyone or the second-order belief that a given moral claim is false for everyone. Finally, we use "first-order" to refer to beliefs that a given moral claim is true or that it is false, and "second-order" to refer to beliefs about whether the truth or falsity of a given moral claim is relative or universal.

universal specifically predicts a high degree of consensus. As such, it is rational to use information indicating high consensus concerning the truth of a moral claim as evidence for universalism with regard to that claim. In the case of low consensus, the universalist hypothesis is penalized for failing to adequately fit the data, while the relativist hypothesis fits the data. So, it is rational to use information indicating low consensus concerning the truth of a moral claim as evidence for relativism with regard to that claim.

Nichols's conclusion hinges on a number of substantive assumptions, however, as he notes. This includes that people are reasonably good at assessing the truth-value of first-order claims and that their first-order beliefs are formed largely independently of other people's first-order beliefs. If either of these failed to hold, then high consensus would not be good evidence for universalism and low consensus would not be good evidence for relativism. Our focus in this paper is on the independence assumption. We argue that there is a potential problem here.

To introduce the problem, let's begin with a simpler case. If people conform their first-order belief about a moral claim to the perceived consensus, independence would clearly be threatened: people's first-order beliefs would be influenced by the first-order beliefs of others. We contend that the same issue arises if people instead call on the perceived consensus to form their second-order belief about a claim and then bring their first-order belief into line with their second-order belief if they are inconsistent. To illustrate, suppose that a person believes that a moral claim is false and that by observing others she comes to believe that a strong majority disagree with her, holding that the claim is true. If this person were to conform her first-order belief to the perceived consensus, coming to agree that the claim is true, her updated belief would not be independent of the beliefs of others. Now suppose that instead of directly updating her first-order belief, she instead calls on the consensus information to form the second-order universalist belief that the claim is true for everyone. Forming this belief on its own would not

alter the independence of her first-order belief that the claim is false. But, her first-order belief would now be inconsistent with her second-order belief. She would hold both that the claim is false *and* that the claim is true for everyone. Faced with this inconsistency, if the person then updated her first-order belief to bring it in line with her second-order belief, her updated belief would no longer be independent of the beliefs of others. While our hypothetical subject would not be *directly* conforming her belief to the perceived consensus, she would nonetheless be doing so *indirectly*, and the result would be the same—in each case she adopts the consensus belief.

Our worry is that due to this potential failure of independence, if people employ the strategy Nichols suggests, then over time their first-order beliefs will tend to converge, and if this happens then over time second-order belief in relativism would tend to decrease as the population comes to converge on either the belief that the claim is universally true or the belief that the claim is universally false. If this worry is accurate, then lay metaethical beliefs would not clearly be vindicated: whether consensus information is good evidence for metaethical beliefs would depend, among other things, on whether the process of calling on consensus information to form second-order beliefs had notably compromised the independence of the first-order beliefs called on in assessing the consensus.

To better determine the plausibility of the above worry we ran a series of computer simulations. The simulations suggest that it is in fact a live worry, but that whether independence is sufficiently undermined to cast doubt on Nichols's process vindication argument depends on an array of factors concerning how people evaluate consensus and the starting distribution of first-order beliefs within and across populations. Here is how we will proceed. In Section 1, we consider Nichols's process vindication argument in more detail. In Section 2, we discuss Nichols's substantive assumptions, focusing on independence. And, in Section 3, we present the results of the simulations we ran.

1. Process Vindication Argument

Nichols's process vindication argument attempts to vindicate lay metaethical beliefs by showing that they are based on a rational process. The general schema for a process vindication argument runs as follows: "If process Q is a rational basis for coming to believe that P, then insofar as S believes that P as a result of process Q, S is (prima facie pro tanto) rational in believing that P." (2019a, 104). Nichols then applies this to lay metaethical beliefs, offering a statistical learning account of how people arrive at these beliefs, arguing that this process works by assessing consensus surrounding the corresponding first-order beliefs, and presenting evidence that people are in fact sensitive to consensus information.

To make it clear what is meant by "rational," Nichols refers to a well-known statement by Edward Stein (1996, 4)—"to be rational is to reason in accordance with principles of reasoning that are based on rules of logic, probability theory and so forth." Nichols's first task, then, is to make the case that the process of calling on consensus information concerning first-order beliefs to arrive at second-order beliefs accords with appropriate principles of reasoning. Calling on Marr's (1982) account of the levels of analysis for processes, Nichols focuses the discussion on the computational level, arguing that a process is *computationally rational* when "*what* the process computes—the input-output profile of the process—corresponds to the function of the relevant logico-probabilistic rules" (2019a, 106). With regard to computational rationality, Nichols then further distinguishes between *global computational rationality* and *local computational rationality*. A globally computationally rational process is one which conforms to rules of logic and probability theory in all cases, whereas a process may be locally computationally rational even if it only conforms to such rules in some restricted range of cases.

Nichols's first key claim, then, is that using consensus information as evidence for metaethical beliefs is locally computationally rational. This limited claim is the main focus of his

process vindication argument, although as discussed in greater detail in the next section, he takes it that this "points towards a possible vindication for the belief in [universalism]" (2019a, 99). To make the case that the process at issue is locally computationally rational, Nichols begins by considering an analogous process involving disease diagnosis. Suppose you are a physician tasked with determining which of two diseases is present in your community. You have two hypotheses, each with a prior probability of 0.5:

H1: The disease is M, which produces a high fever and no other symptoms.

H2: The disease is D, which produces either a high fever or a sore throat, but never both.

Suppose that ten patients come to see you, five of which have a high fever, while the other five have a sore throat.[3] **H1** is a poor fit for the data because it would require dismissing half of the data as noise. On the other hand, **H2** is a perfect fit for the data because it can explain the symptoms of all ten patients. Now suppose an alternative distribution of symptoms, with nine patients having a high fever and just one having a sore throat. Again, **H2** perfectly fits the data, since it can explain all of the data points just as before. But **H2** will fit any distribution of these two symptoms—it is a highly flexible hypothesis. Nichols argues that there is a tradeoff, here, between fit and flexibility: the degree to which a hypothesis fits the data counts in favor of it, but the degree to which it is flexible counts against it. While **H2** fits the data perfectly, it is penalized for its extreme flexibility. And since **H1** also fits the data well in the second case, while being far less flexible than **H2**, Nichols argues that **H1** is to be favored in this situation. And he contends that deciding between hypotheses in this way is rational because it accords with a principle of reasoning, namely Bayesian probability theory.

[3] The numbers for this example are taken from an earlier version of Nichols (2019). Ayars and Nichols (2019) use the same example but shift the numbers for this case to six patients having a high fever and four having a sore throat, although this doesn't make a notable difference for our discussion. The published version of Nichols (2019a) shifts the sample to 20 patients rather than ten. The implications of this difference in sample size are explored in Section 3.

Nichols then shows how this same kind of reasoning can be applied to the process of arriving at second-order moral beliefs on the basis of information about the distribution of first-order beliefs. Consider two hypotheses that could explain people's first-order beliefs about a given moral claim, each with a prior probability of 0.5:

> **Universalism:** There is a single moral fact about whether φ-ing is wrong.
>
> **Relativism:** There is no single moral fact about whether φ-ing is wrong; rather, whether φ-ing is wrong is relative to context or culture.

In line with the above example, suppose that you survey ten people and find that five of them have the first-order belief that φ-ing is wrong and that the other five have the first-order belief that φ-ing is not wrong. On the assumption that people are good at assessing moral facts, **Universalism** is a poor fit for the data because it would require dismissing half of the data as noise. On the other hand, **Relativism** is a perfect fit for the data, explaining the beliefs of all ten people. Now suppose an alternative distribution of beliefs, with nine of the people holding the first-order belief that φ-ing is wrong and only one holding the first-order belief that φ-ing is not wrong. Again, **Relativism** perfectly fits the data, explaining all of the data points just as before. But, of course, **Relativism** will fit any distribution of these two beliefs. Since **Universalism** also fits the data well, and since it is far less flexible than **Relativism**, Nichols contends that **Universalism** is to be favored in this situation, just as **H1** was favored in the previous example. Based on this Nichols argues that if people call on consensus information to arrive at beliefs about universalism and relativism, this process would be rational (other things being equal) because it accords with principles of reasoning (Bayesian probability theory). And he restricts this to a local rationality claim, since the process might not accord with principles of reasoning for all cases.

Nichols's second task is to make the case that people are in fact sensitive to consensus information in arriving at second-order moral beliefs. For this he points to previous empirical

evidence suggesting both a correlation between consensus information and second-order beliefs (Goodwin & Darley 2008) and that perceived consensus can cause second-order beliefs (Goodwin & Darley 2012), as well as adding to this body of evidence (Ayars & Nichols 2019). Since our focus in this paper is on the implications of the process of calling on consensus information to form second-order beliefs, we will not examine this evidence further here, but will proceed under the assumption that people do in fact employ such a process.

2. Independence Assumption

Nichols rightly notes that the rationality of the inference from consensus information about first-order beliefs in a claim to universalism or relativism about that claim hinges on a number of substantive assumptions. It is assumed that the agent decides between just these two hypotheses and that the priors for each are equal. Further, it is assumed that there are first-order moral facts, that people are generally good at assessing them, and that people believe that other people are generally good at assessing them. Finally, people must assess the moral facts independently of one another and must believe that people assess the moral facts independently. While each of these assumptions might be challenged, we will focus on the last two—independence and belief in independence.

Nichols makes clear at several points that his argument hinges on people by-and-large arriving at first-order beliefs independently. For instance, Ayars and Nichols (2019, 7) write:

> individuals' judgments must be, to some significant extent, independent. It can't be that everyone is just blindly copying the opinion of one person. The individuals who make up the distribution of responses must provide independent evaluative data points. We hope that future work will investigate these matters systematically, but for present purposes we will rely on the assumption that people do expect that others are good at tracking evaluative facts and that individuals' evaluations are largely independent. With these assumptions in place, it is rational to use consensus as evidence regarding universalism and relativism. Indeed, insofar as these assumptions are plausible, it is rational for us—not just our participants—to use consensus information as evidence regarding universalism and relativism.

Independence is also discussed in Nichols (2019a, 4.5.2.3). There he more clearly distinguishes between the two assumptions noted above—that people's first-order beliefs are (to some significant extent) independent and that people believe that people's first-order beliefs are independent. Relatively little is said about the former, beyond noting that "individuals must provide independent evaluative datapoints." The latter is instead treated as the "primary question" and is discussed at more length.

Presumably the claim that the primary question is *belief in independence*, rather than *independence itself*, reflects that Nichols's main concern is with arguing for the local computational rationality of the process. There are two ways in which this might be defended if belief in independence holds but independence itself does not. The first defense focuses on how the process could remain computationally rational. Here the concern is with the input-output profile of the process. As such, if belief in independence is taken to be an input to the process Nichols is concerned with, the process could be computationally rational even if the input turns out to be false or the process generating it irrational. The truth of the input and/or the rationality of the process producing it would be beside the point for assessing the rationality of the process that input feeds into. The second defense concerns just how *local* the computational rationality of the process is. Nichols holds that the process at issue will be computationally rational for some inputs but not others. While he does not specify the relevant range of inputs, this might be taken to exclude those claims where independence does not hold. In other words, the process at issue would be computationally rational for at most those inputs where the independence assumption holds. On this defense, justifying the independence assumption would play a key role in determining the bounds on the local computational rationality of the process.

Here it should be noted that Nichols's ultimate goal goes beyond arguing for the local computational rationality of the process at issue; he also holds that this can point toward a

possible vindication of people's second-order beliefs. If the process of forming beliefs based on consensus information is computationally rational but takes in key assumptions (such as independence) that are not themselves justified, however, then the rationality of the process would not seem to provide much vindication for metaethical beliefs. And if the process of forming beliefs based on consensus is only rational in circumstances that seldom hold when people employ the process, then the rationality of the process would do little to provide a general vindication of metaethical beliefs. Further, as noted above, Nichols treats independence as a condition on the rationality of using consensus information as evidence for second-order beliefs, even if he takes belief in independence to be the primary question. Nonetheless, if the assumption of independence is instead sequestered off from the process at issue in some way, the present work should be taken as exploring a further question that arises in vindicating lay metaethical beliefs.

Our key question, then, is whether independence holds if people call on consensus information in coming to hold second-order beliefs. Unsurprisingly, the notion of independence at play here is unclear. As Dietrich and Spiekermann (2013, 655) note, "the concept of opinion independence is not well understood, even though it is crucial in social epistemology." They go on to distinguish four independence assumptions and show that two of them can be used in jury theorems (i.e., formal "wisdom of the crowd" arguments). Whether or not one of these assumptions is also appropriate for Nichols's process vindication argument is complicated by the fact that the process he proposes is not a straightforward use of the wisdom of the crowd, since people would be calling on the first-order beliefs of others to form a second-order belief and not (directly) for purposes of determining what the correct first-order belief is.[4] Further, it is the pattern of first-order beliefs from multiple members of the crowd, not their individual first-order

[4] Insofar as the ultimate aim here is to vindicate lay metaethical beliefs, however, Nichols's process vindication argument, if accurate, would feed into a wisdom of the crowd argument with regard to second-order beliefs.

beliefs themselves, that serve as evidence for inferring a second-order belief. Fortunately, our worry with regard to independence is not specific to any of the concepts that Dietrich and Spiekermann distinguish. Our worry is that people's first-order beliefs might *cause* other people's first-order beliefs, which would be a violation of independence on any plausibly applicable concept.

The first reason for worry is that people likely call on other people's first-order beliefs in forming their own first-order beliefs. As Nichols (2019a, 117) notes in arguing for another substantive assumption (that people expect others to be good at first-order moral judgments), there is extensive evidence that this occurs: "Work on conformity indicates that people will use other's actions and normative judgments as evidence about the right thing to do." Such work raises two forms of "harmful opinion dependence" that Dietrich and Spiekermann (656) note can undermine the wisdom of the crowd—information cascades and systematic biases. We'll focus on the former, as it fits in with our second reason for worry.

The second reason for worry, and the one we will focus on, is that there is a theoretical connection between second-order beliefs and first-order beliefs. Ayars and Nichols (2019, fn11) note that first-order claims can have implications for second-order claims. But things also go the other direction: if it is a second-order fact that there is a single first-order fact—that φ-ing is wrong, say—then the first-order thesis that φ-ing is wrong obviously follows. Focusing on beliefs, it seems that if you arrive at the second-order belief that there is a single moral fact that φ-ing is wrong, then consistency compels you to also adopt the first-order belief that φ-ing is wrong. And if you arrive at the second-order belief that there is a single moral fact that φ-ing is *not* wrong, then consistency compels you to also adopt the first-order belief that φ-ing is *not* wrong. The result is that if people call on consensus information to form their second-order beliefs it seems that rationally this should affect their corresponding first-order beliefs. But since

this would be carried out on the basis of (we are assuming fairly accurate) evaluations of the first-order beliefs of other people, the revised first-order beliefs would not be independent of other people's first-order beliefs.

Perhaps it could be argued that people do not in fact update their first-order beliefs to make them consistent with their second-order beliefs. One possibility is that people are generally willing to tolerate this type of internal inconsistency. While we find this unlikely, it is an empirical assumption that would need testing. A second possibility is that people typically only call on consensus information to arrive at second-order beliefs that are consistent with the first-order beliefs they already hold. Empirical work on information cascade arguably pushes against this possibility (Ziegelmeyer et al. 2010). If people nonetheless typically only call on consensus information to arrive at second-order moral beliefs that are consistent with their first-order beliefs, however, this type of "application bias" would seem to challenge the rationality of the overall process. It might be argued that the situations in which a process is employed is external to the process itself, such that bias in when the process is called on would not compromise the computational rationality of the process. This would further restrict the claim that lay metaethical beliefs are partly based on rationally appropriate processes, however, and the selective use of the process in this way would do little to vindicate lay metaethical beliefs. For the remainder of this paper, we'll set these worries aside to focus on the implications for the independence assumption if people do in fact typically conform their first-order beliefs to the consensus information in the way outlined above.

Making this assumption, is the theoretical failure of independence noted enough to undermine the assumption that first-order beliefs are "to some significant extent" independent? That is difficult to say. One reason is that it is unclear what the required degree of independence is. Further, it is plausible that this will depend on the details of the model, including potentially

how people sample consensus data from the population, what the threshold is for adopting a second-order belief, and whether and to what extent people remember previous samples of consensus data, among other factors. Nichols does not attempt to offer a detailed model for how we arrive at second-order moral beliefs, but simply argues that "using consensus as evidence concerning [universality] is rational" (2019a, 110) and illustrates this with the simple one-off cases described above. As such, we'll take these examples as a guide in the simulations described in the next section.

Like Nichols, we will not attempt to offer a detailed model here. Rather, our goal is to begin to explore what further assumptions are needed to avoid a problematic failure of independence. More specifically, our concern is with putting some initial bounds on when belief in relativism will be stable if people call on consensus information in forming their second-order beliefs and maintain consistency between their second-order beliefs and their first-order beliefs. Toward this we ran a series of computer simulations to see how the beliefs of a population would change over time on different assumptions.

3. Simulations

We begin with a basic set of simulations to illustrate that relativist beliefs are potentially unstable if consensus information is used to form second-order beliefs and first-order beliefs are then brought into line with second-order beliefs. We then progressively add some complications to the models.[5]

[5] All simulations were run in R. There are several benefits to using R in the present context. Perhaps most importantly is that many experimental philosophers are familiar with R and the only book-length guide to the practice of experimental philosophy currently available (Sytsma & Livengood 2016) uses R as its preferred statistical program. Chapter 10 of that volume provides a general introduction to the use of R.

3.1 Initial Simulations

In our first set of simulations we modeled a population of 1000 people who initially have either the first-order belief φ or the first-order belief not-φ and no second-order belief about φ. The initial distribution for first-order beliefs is determined randomly with each person having a 50% chance to hold φ and a 50% chance to hold not-φ. The simulation is then run over 10k timesteps.[6] At each timestep a random 1% of the population checks the first-order beliefs of a random 1% of the population, excluding themselves.[7] This gives a sample size of 10 for each evaluation, which corresponds with the example given in an earlier version of Nichols (2019a). If first-order beliefs that φ among the sample are greater than a set consensus threshold—either 70%, 80%, or 90%—then that person adopts the second-order belief of universalism and brings their own first-order belief in line with that judgment (i.e., if they hold not-φ they switch to φ).[8] Similarly, if first-order beliefs that not-φ among the sample are greater than the consensus threshold, that person adopts the second-order belief of universalism and changes their first-order belief accordingly. Finally, if neither of these holds, the person adopts the second-order belief of relativism and their first-order belief remains unchanged. In this initial set of simulations people

[6] It might be argued that this is longer than Nichols needs to worry about—that if relativism is stable over an initial relatively short period of time, that is sufficient for the rationality of the process. In addition, it might be argued that people do not evaluate that many first-order beliefs. The present parameters would have the average person assessing the consensus 100 times over the run of a simulation. Over the course of a lifetime, this doesn't strike us as unrealistic, although it would likely depend on exactly what the moral claim at issue is. Further, part of the issue here is that people wouldn't know where along this timespan they are. Put another way, it wouldn't be reasonable for people to assume that they are near the initiation point when they call on consensus data. Finally, while we will not model people dying off or being born in our simulations, the real-world process of forming moral beliefs can be thought of as extending well beyond the life of an individual. And if children are likely to learn their initial moral beliefs from their family, then we might want to think about this in terms of extended lineages.

[7] This follows the examples given in Section 1. It might be thought, however, that it would be better to have people always include their own first-order belief in assessing the consensus information. This would essentially introduce a (fairly small) element of bias toward universalism about their own first-order belief and against universalism about the opposite of their own first-order belief. With regard to the implications of second-order belief on first-order belief, this would in effect serve to slightly raise that person's consensus threshold: a higher consensus against the person's first-order belief would be needed for the person to change their own first-order belief. Since we already vary the consensus threshold that people employ, this can be thought of as already being captured in our simulations.

[8] Nichols does not discuss where the threshold between high and low consensus should be. The examples he provides, and those used in the empirical literature, aim to give clear instances of consensus or failure of consensus, not to explore what threshold people use. As such, we chose to run these simulations using several different consensus thresholds. And as we will see, what threshold people actually use will matter for the stability of relativism over time.

had no memory for past evaluations and had perfect access to the first-order beliefs of a given sample. 100 simulations were run for each consensus threshold (Figure 1).

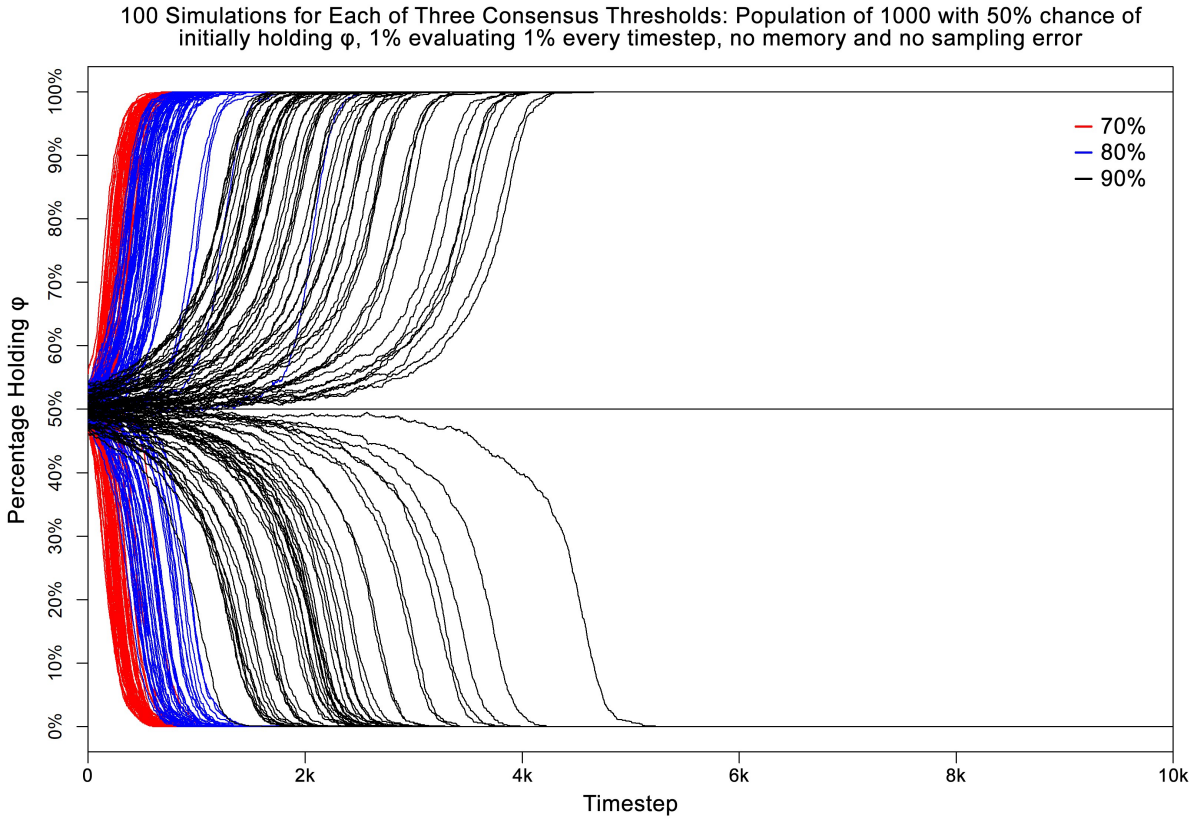
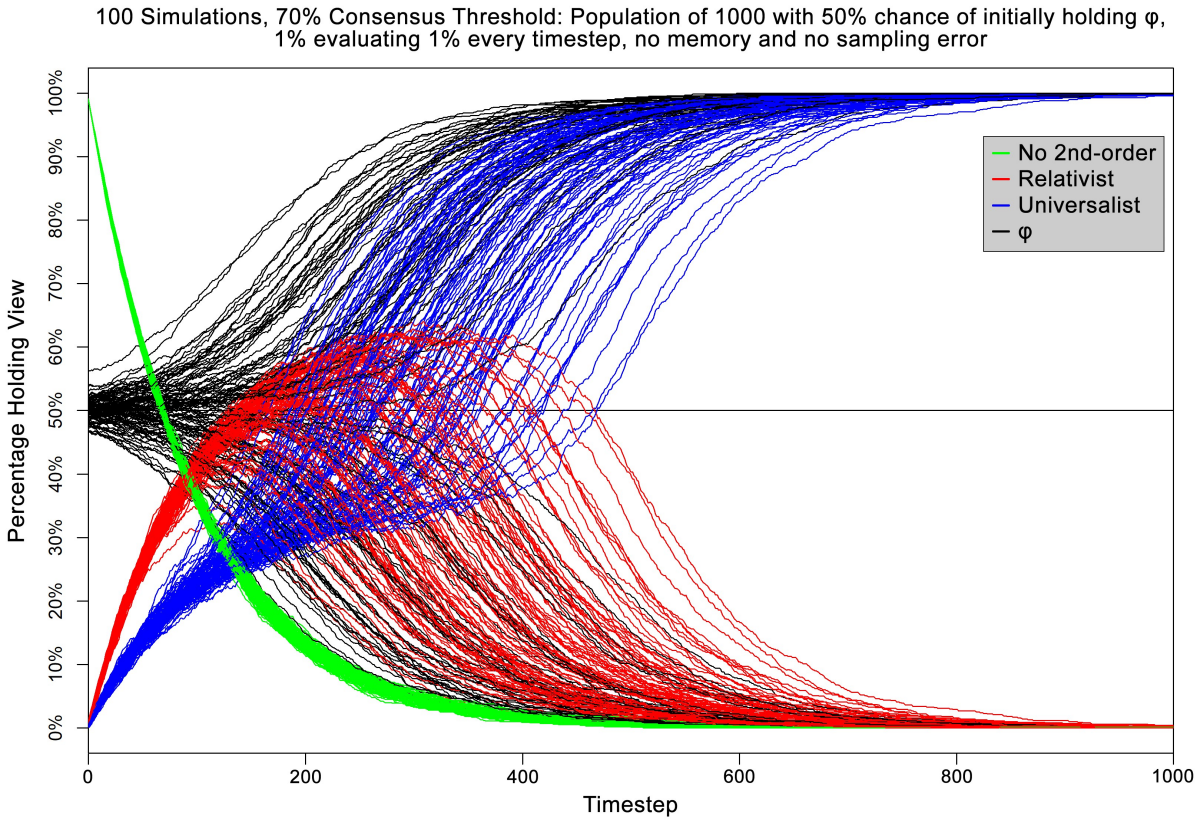


Figure 1: Simulations 1, Base

What we find is that across the simulations the first-order beliefs of the population shift to either 100% φ or 100% not-φ. It follows that the second-order beliefs of the population shift to 100% universalist. Despite roughly half the population starting with each first-order belief, the population does not end up with relativist views. This is our first important finding: the process at issue in Nichols's process vindication argument can be susceptible to problematic breakdowns in independence due to the effect of second-order beliefs on first-order beliefs.

To better understand how metaethical positions are changing over time, we ran a second set of simulations using the same parameters as Simulations 1 but also plotting second-order

beliefs. To better visualize the change in positions, we plotted a truncated number of timesteps (Figure 2). What we find is that that there is an initial upswing in relativist beliefs, followed by a rise in universalist beliefs and a corresponding fall in relativist beliefs. Not surprisingly, the degree of the rise in relativist beliefs and the time before those beliefs fall depends on the consensus threshold employed, with relativism having a better run the higher the consensus threshold.



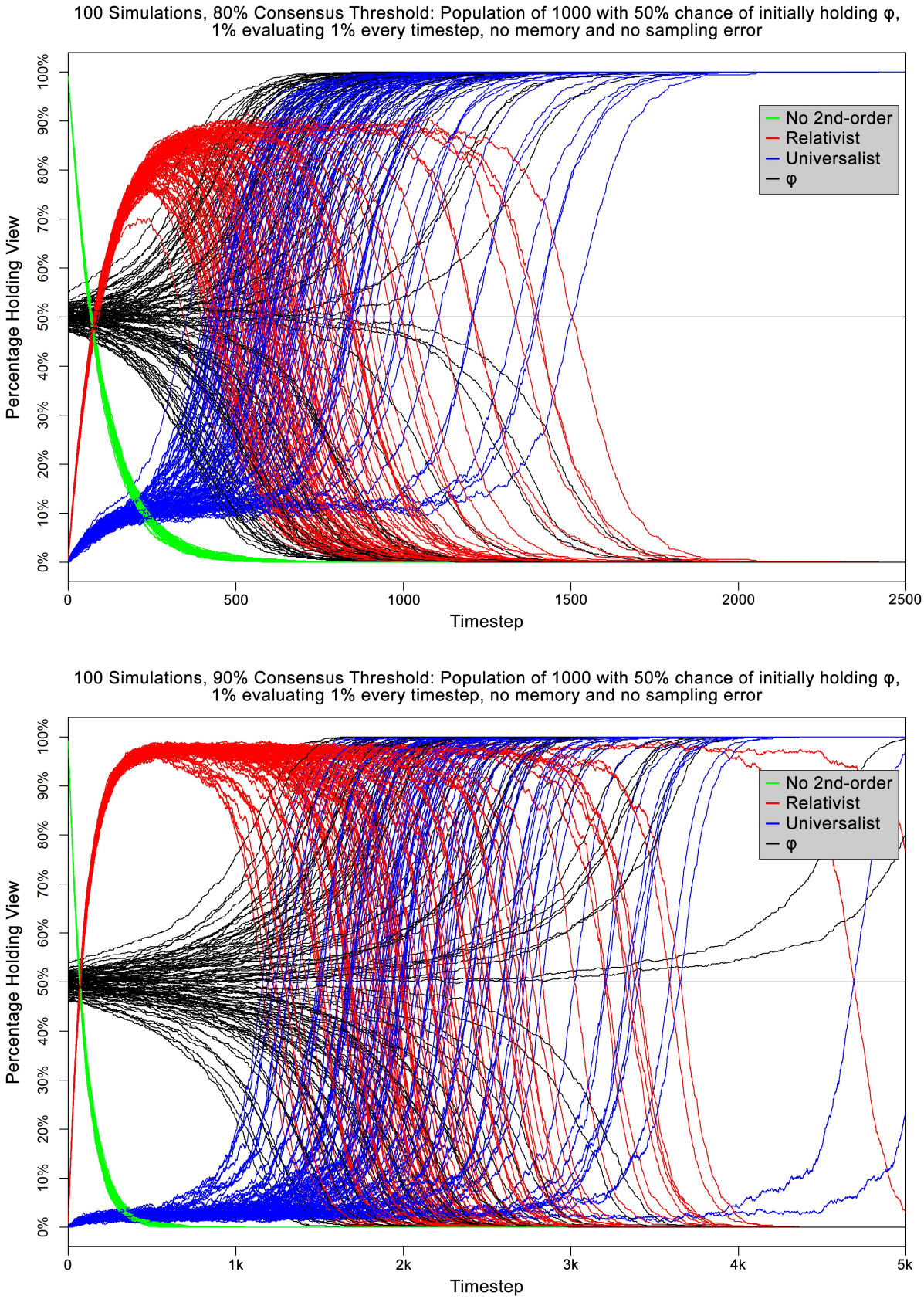


Figure 2: Simulations 2, Varying Consensus Threshold

While an earlier version of Nichols (2019a) and Ayars & Nichols (2019) give examples where the sample size is ten, the published version of Nichols (2019a) and Nichols (2019b) give the same example with a sample size of 20.[9] We would expect relativism to become more stable as sample sizes increase, since the probability of drawing an initial sample that exceeds the thresholds will decrease with sample size. As such, we ran a third set of simulations using the same parameters as before but with a sample size of 20, increasing the number of evaluations per timestep to match, and including consensus thresholds of 75% and 85%. For purposes of comparison, we also ran a set of simulations using a sample size of 50 and increasing the number of evaluations per timestep to match. The results are shown in Figure 3. We find that the simulations using a sample size of 20 are comparable to the initial set, although convergence is delayed. Illustrating the importance of sample size, using a sample size of 50, no convergence was seen through 10k timesteps for the highest two thresholds (85%, 90%). We can draw two further conclusions: whether the process at issue is likely to show a problematic breakdown of independence over time is sensitive to the number of first-order beliefs people assess before forming a second-order belief and is sensitive to how extreme the threshold is.

[9] It is unclear what people's actual sampling behavior looks like, including whether a sample size of ten or 20 is more realistic for our simulations. There is evidence, however, from Nichols, van Roojen, & Murray (forthcoming) indicating that sample size affects consensus informed beliefs about the universality of jokes, with smaller samples leading to less universalist beliefs. Whether or not people are sensitive to sample size for moral questions needs to be tested.

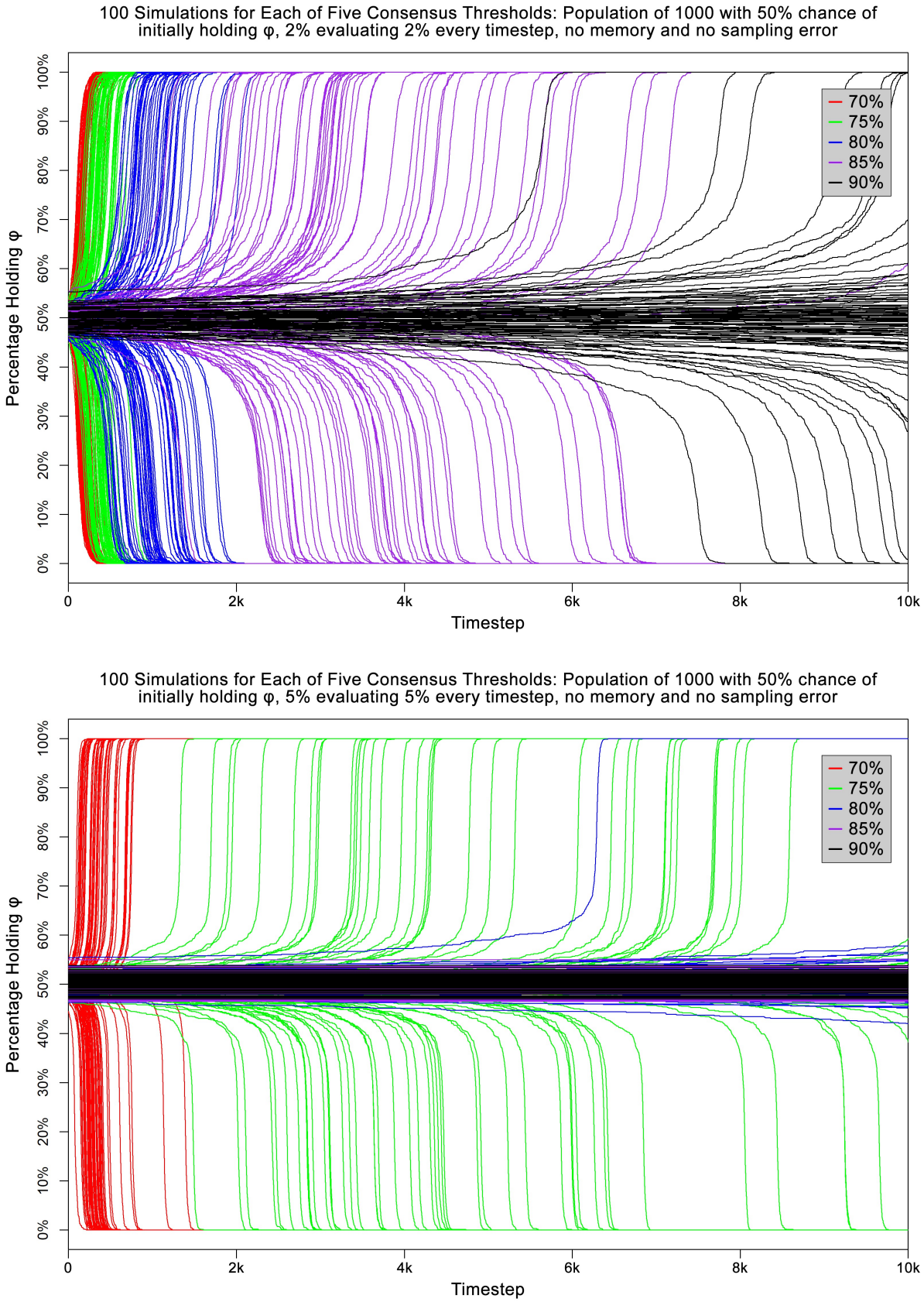
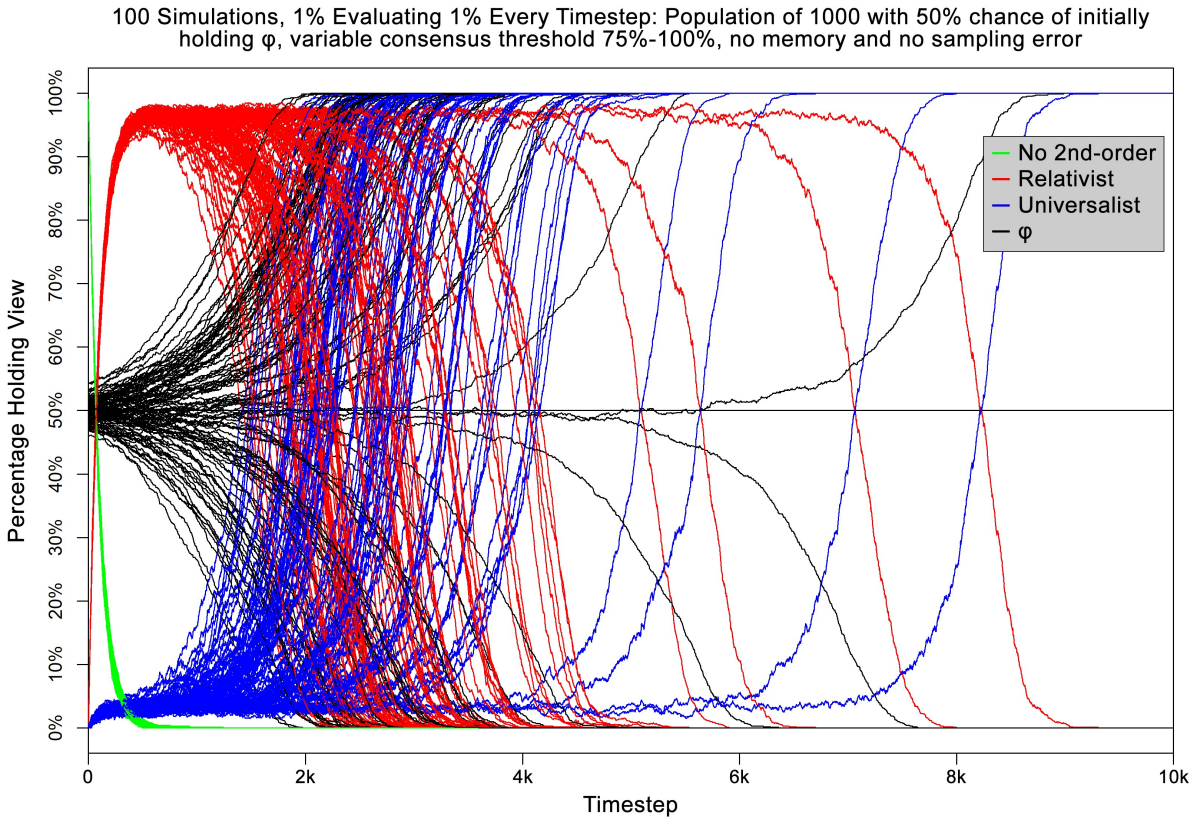


Figure 3: Simulations 3, Varying Sample Size

3.2 Variable Consensus Threshold

Obviously, there are great number of simplifying assumptions made in the above simulations, as in any such simulation work, and many of these are likely to be relevant to assessing the stability of relativist beliefs. One simplification is that every person used the same consensus threshold in each simulation, although we varied the threshold across sets of simulations. As noted above, the literature doesn't provide clear guidance here. One of Nichols's (2019a) assumptions is that the two hypotheses each have a prior probability of 0.5, which suggests an equal partitioning. A lower threshold for consensus of 75% would equally partition the space between universalist and relativist views and is consistent with what has been used in the empirical literature. In our fourth set of simulations, we used the same parameters as Simulations 1 (sample size 10) and the first set in Simulations 3 (sample size 20), but this time randomly assigned each person a consensus threshold between 75% and 100% (Figure 4). Again, we find that first-order beliefs converge, or are converging, on either 100% φ or 100% not-φ, with a corresponding convergence to universalism. As before, we find the same basic result whether using a sample size of either 1% or 2% of the population. As such, to ease computation time we will use the former in subsequent simulations.



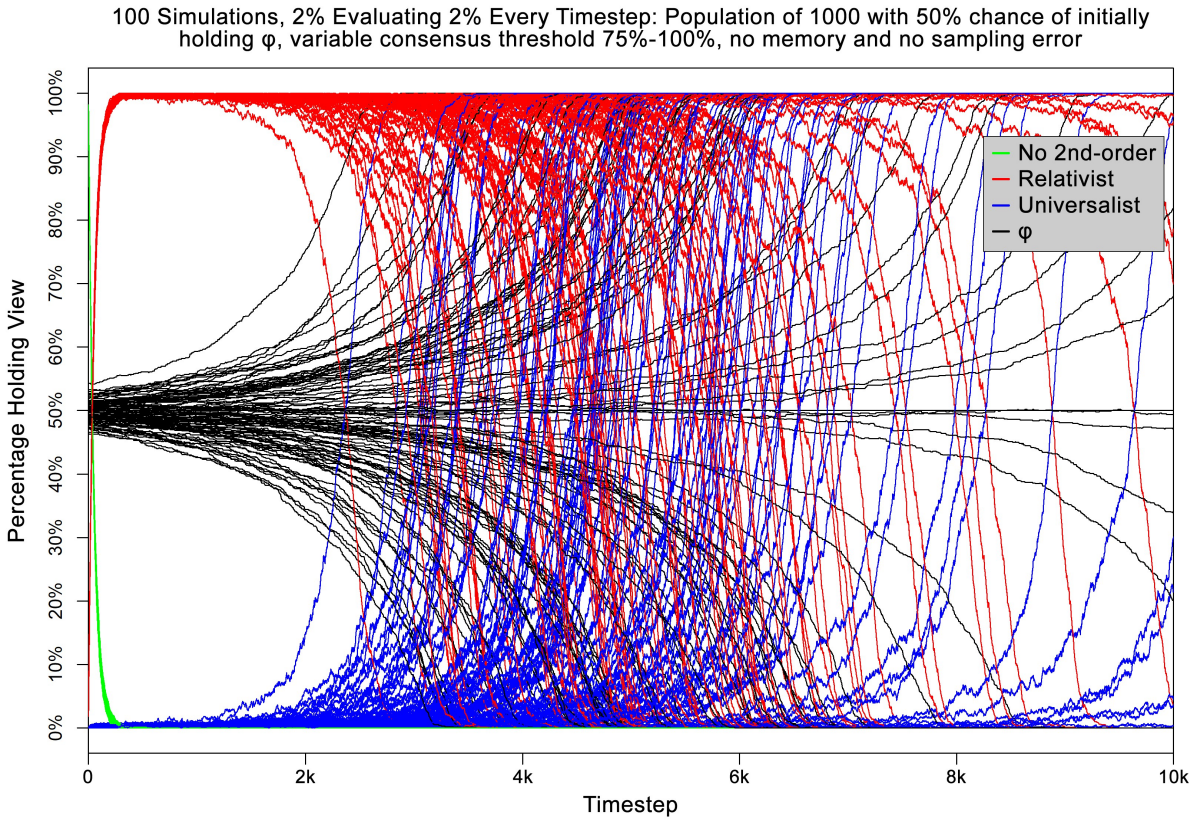
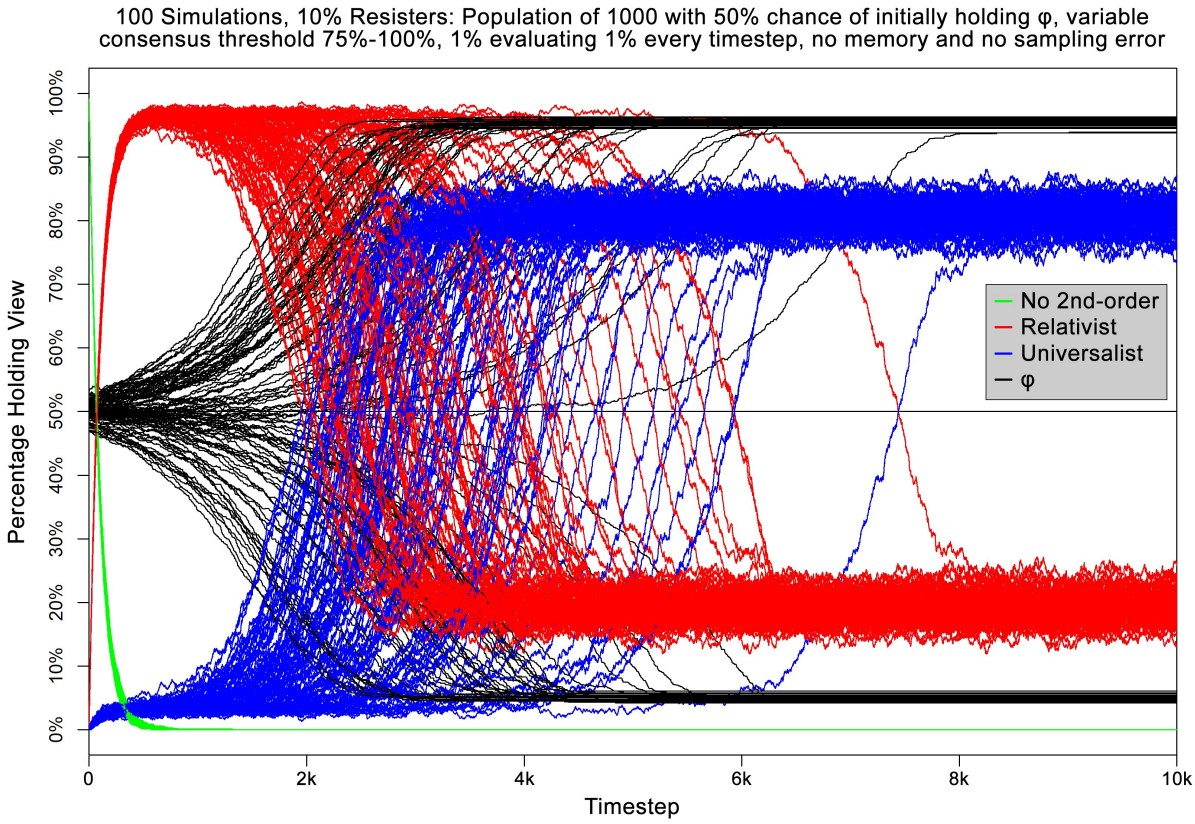
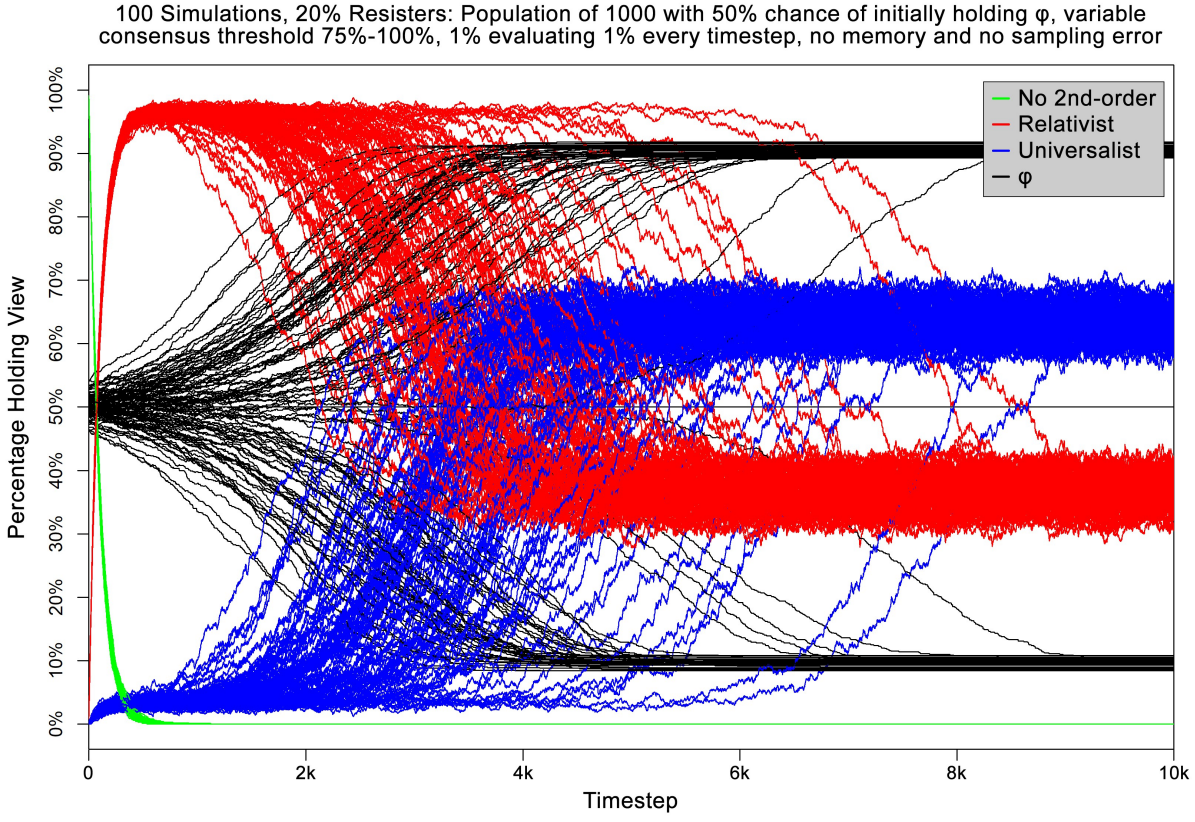


Figure 4: Simulations 4, Variable Consensus Threshold

3.3 Minority Doesn't Change First-order Belief

While we noted in the previous section that we find it unlikely that people in general would tolerate inconsistency between their first-order and second-order beliefs, we find it plausible that some people would do so. To test the effect this would have, we ran a fifth set of simulations using the same configuration as in the first set in Simulations 4, but added a minority who does not change their first-order belief regardless of the consensus information. We ran this for three minority percentages, making the resisters 10%, 20%, or 30% of the total population (Figure 5). Of course, first-order beliefs no longer converge on either 100% φ or 100% not-φ, since roughly 5% or 10% or 15% of the population are resisters who will continue to hold the opposite belief regardless, but the remainder of the populations do converge on one or the other of the two beliefs. Likewise, we find that while the populations do not converge on universalist second-order beliefs, a large percentage shifts to universalism, with the size of the percentage depending on the percentage of the population that resists changing their first-order belief. Thus, our fourth important finding is that the problematic failure of independence can persist even if a notable minority of the population does not change their first-order belief regardless of the consensus.


100 Simulations, 10% Resisters: Population of 1000 with 50% chance of initially holding φ, variable consensus threshold 75%-100%, 1% evaluating 1% every timestep, no memory and no sampling error
Percentage Holding View
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0
2k
4k
6k
8k
10k
Timestep
No 2nd-order
Relativist
Universalist
φ


100 Simulations, 20% Resisters: Population of 1000 with 50% chance of initially holding φ, variable consensus threshold 75%-100%, 1% evaluating 1% every timestep, no memory and no sampling error
Percentage Holding View
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0
2k
4k
6k
8k
10k
Timestep
No 2nd-order
Relativist
Universalist
φ

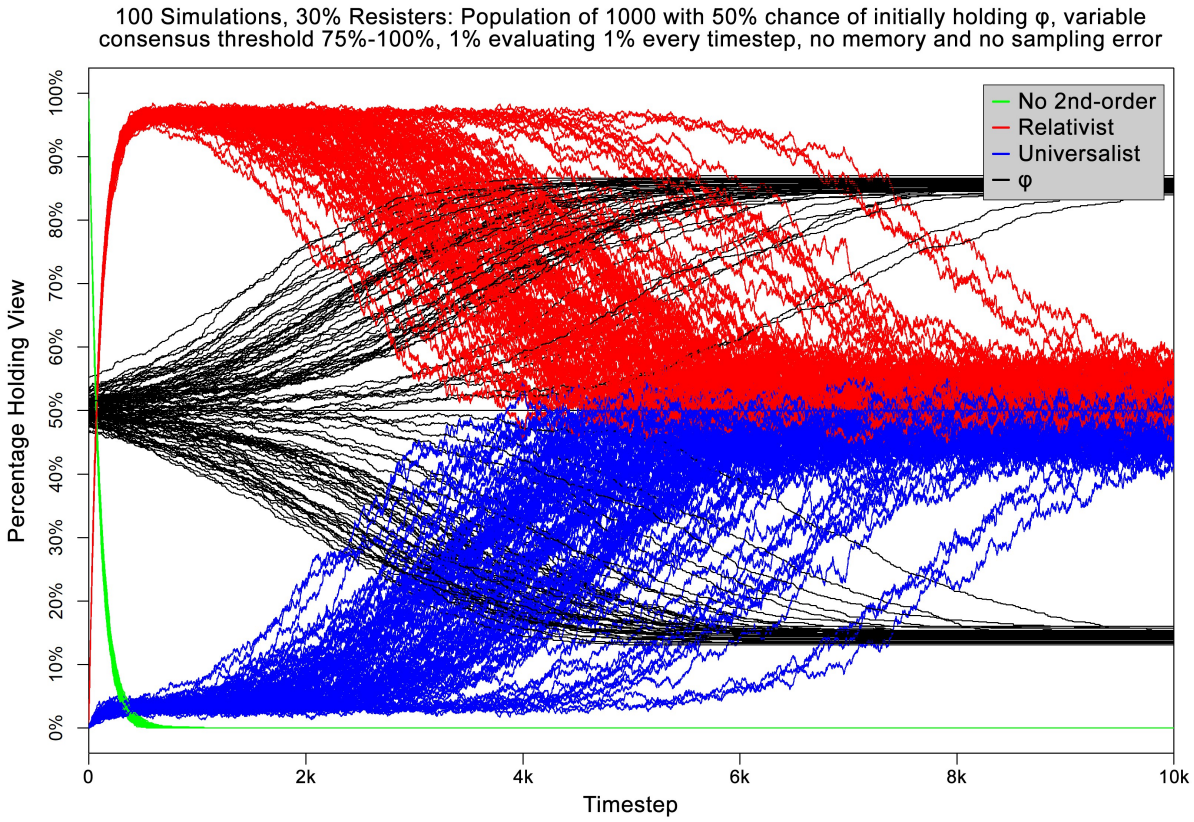


Figure 5: Simulations 5, Resisters

3.4 Perfect Memory

Perhaps the most obvious issue with the previous simulations is that people had no memory: changes to their second-order beliefs were based on just their present sample, not any previous information they had collected. This might be conceptualized as an extreme form of availability bias, with people weighing *only* their most recent evidence. The lack of memory can be expected to reduce the stability of relativism, as it would make it easier for changes in first-order judgments to cascade through the population. To test this, in our sixth set of simulations we went the other direction, having each person remember *all* of their previous evaluations with perfect recall and giving all evaluations equal weight. This might be conceptualized as people showing no availability bias. As seen in Figure 6, this did indeed greatly increase the stability of relativism, with first-order beliefs quickly becoming stable and second-order beliefs quickly

converging on 100% relativist. A single simulation using the same starting population used above and run over 200k timesteps indicates that relativism is indeed stable here.

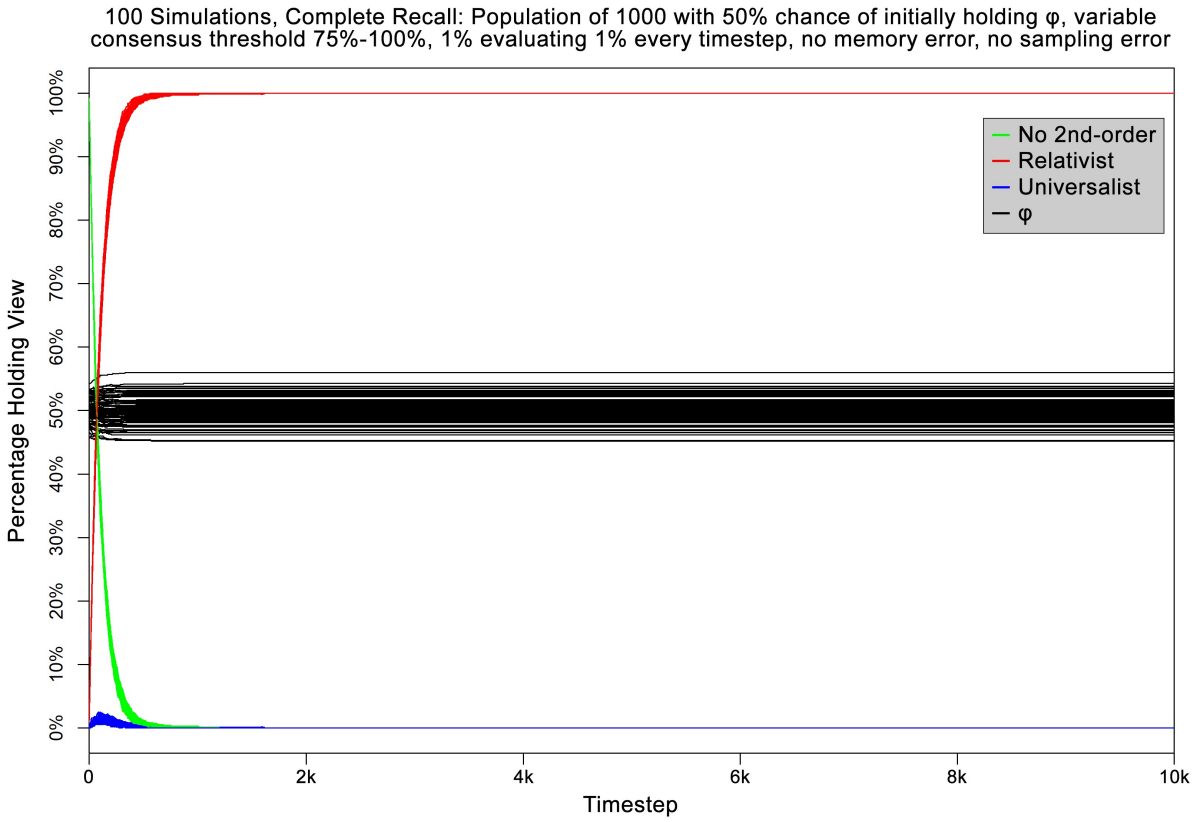


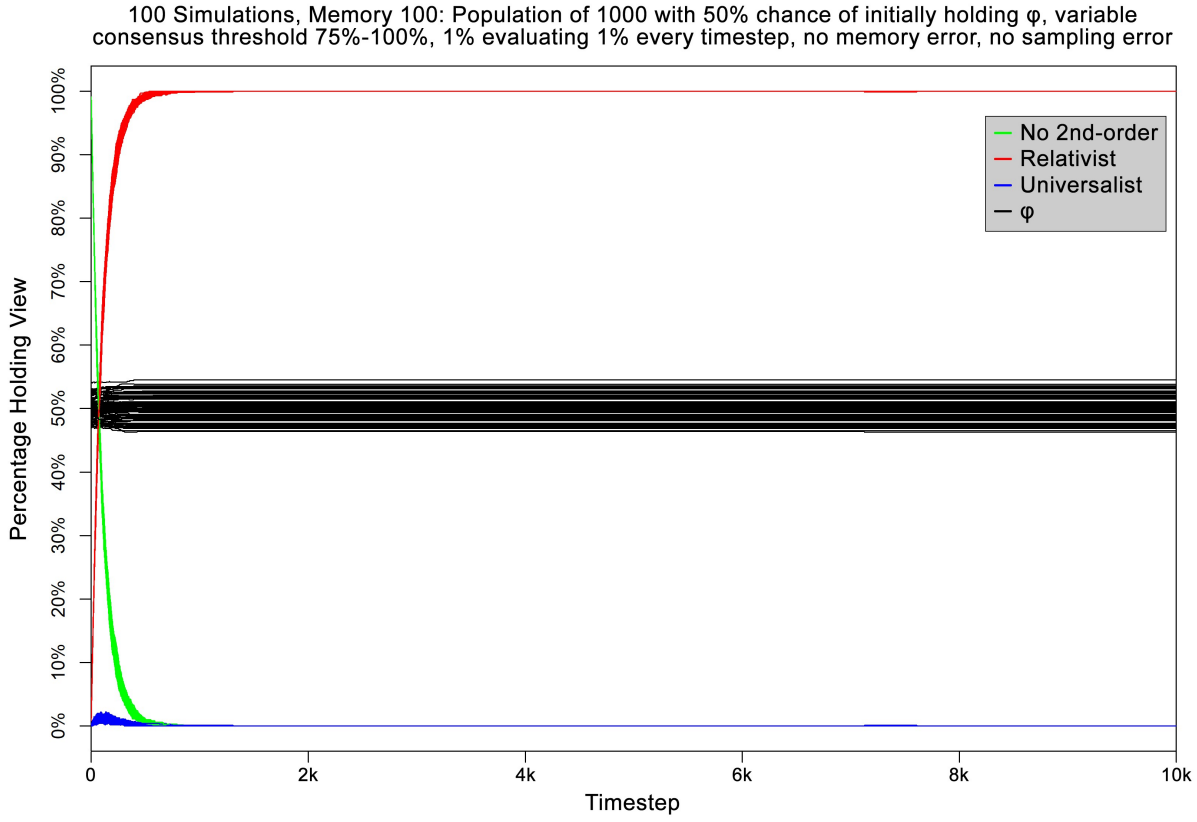
Figure 6: Simulations 6, Complete Recall

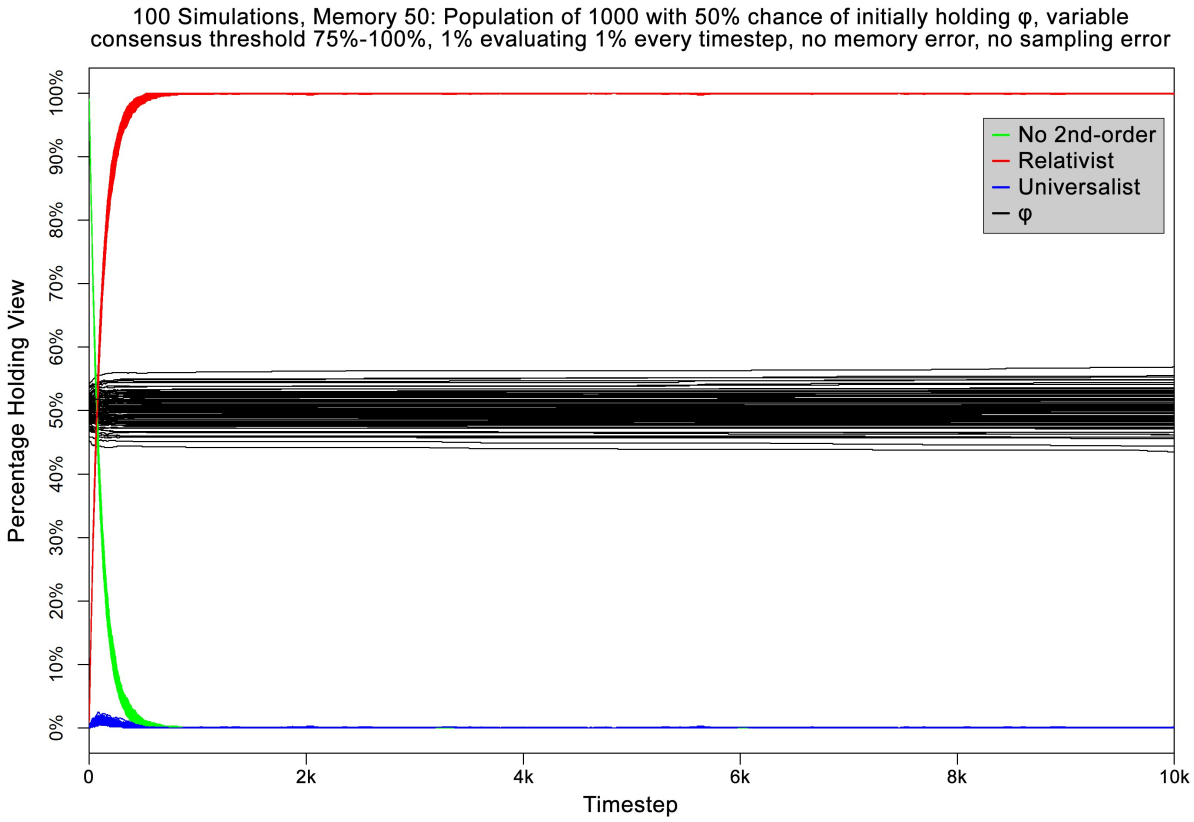
3.5 Partial Recall

Just as it is unrealistic to assume that people would not remember previous evaluations at all, it is also unrealistic to assume that they would have perfect recall for all past evaluations. This was adjusted in our seventh set of simulations, changing the parameters from Simulations 6 to limit memory to the most recent 100, 50, or 25 evaluations. As before, each of these evaluations was remembered with perfect recall and all evaluations were given equal weight. As seen in Figure 7, first-order beliefs quickly become stable over the remainder of the 10k timesteps with a memory of 100 evaluations, and we would expect the same behavior if the memory was increased still

further. In contrast, setting memory to 50 we just start to see a shift in first-order beliefs at the end of the simulations. And setting memory to 25, it is clear that the populations are converging.

To check whether convergence occurred for partial memory, we ran a single simulation using the same starting population and an extended series of timesteps for each of these sets of parameters. The results suggest that first-order beliefs are stable with a memory of 100 evaluations but shift to either 100% φ or 100% not-φ with a memory of 50 evaluations or 25 evaluations. Our fifth important finding is that whether there is a problematic failure of independence and how quickly it occurs is sensitive to the number of evaluations that people are able to remember.





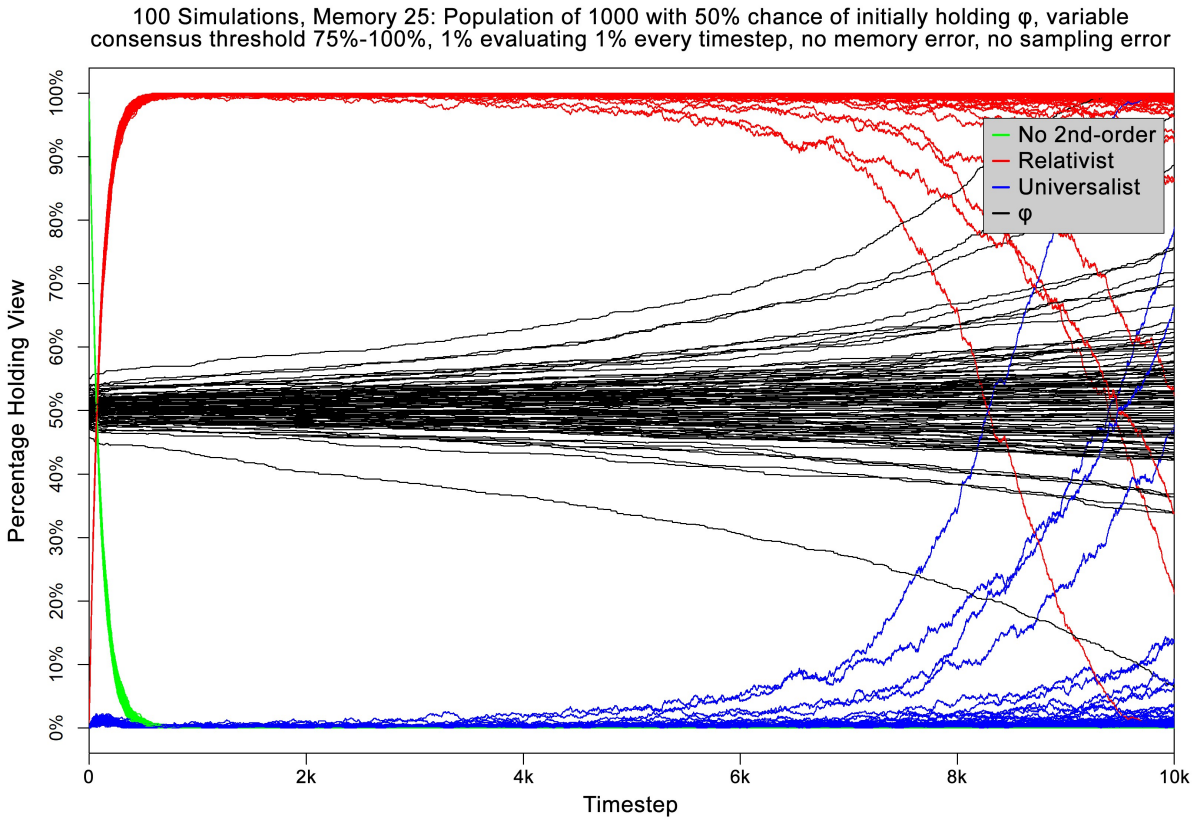


Figure 7: Simulations 7, Partial Recall

3.6 Recency Bias

Just as it is unrealistic to assume that people would have perfect recall for all past evaluations, it is also unrealistic to assume that they would weigh all evaluations equally. It is likely that systematic biases and error will play a role in how people assess consensus, including that more recent evaluations will be more readily available for recall and will be disproportionately weighted as people consider their consensus information. To test this, we added a recency multiplier to the simulations, with the first set of evaluations being considered given a base weighting, then each subsequent set being weighted by X times the previous set. We began by running a single simulation using the same starting population as before with a memory of 100, using one of four recency multipliers (5, 4, 3, 2), and running the simulations over an extended number of timesteps where necessary to observe convergence behavior. Convergence is observed for each of the four recency multipliers, with the time to convergence decreasing as the multiplier increases (5x converging at 4,207 timesteps, 2x at 23,687).

To check that the effect of the recency bias is not specific to the starting population used in the previous simulations, we ran an eighth set of simulations for the 2x recency multiplier using 100 random starting populations run over 30k timesteps (Figure 8). While only some of the simulations converge on 100% φ or 100% not-φ over this span, that the simulations are converging is clear. To further test the interaction between the recency bias and memory size, we ran a single simulation using the same starting population as before and a 2x recency multiplier, with a memory size of 250, 500, 1000, or all evaluations. We found that all four simulations converged at 23,838 timesteps. Further, the plots were only very slightly different from the plot for a memory size of 100. These simulations lead to our sixth important finding: how people's memory works, in addition to just the extent of their memory, matters for the process at issue; specifically, we find that recency bias effectively serves to truncate memory size.

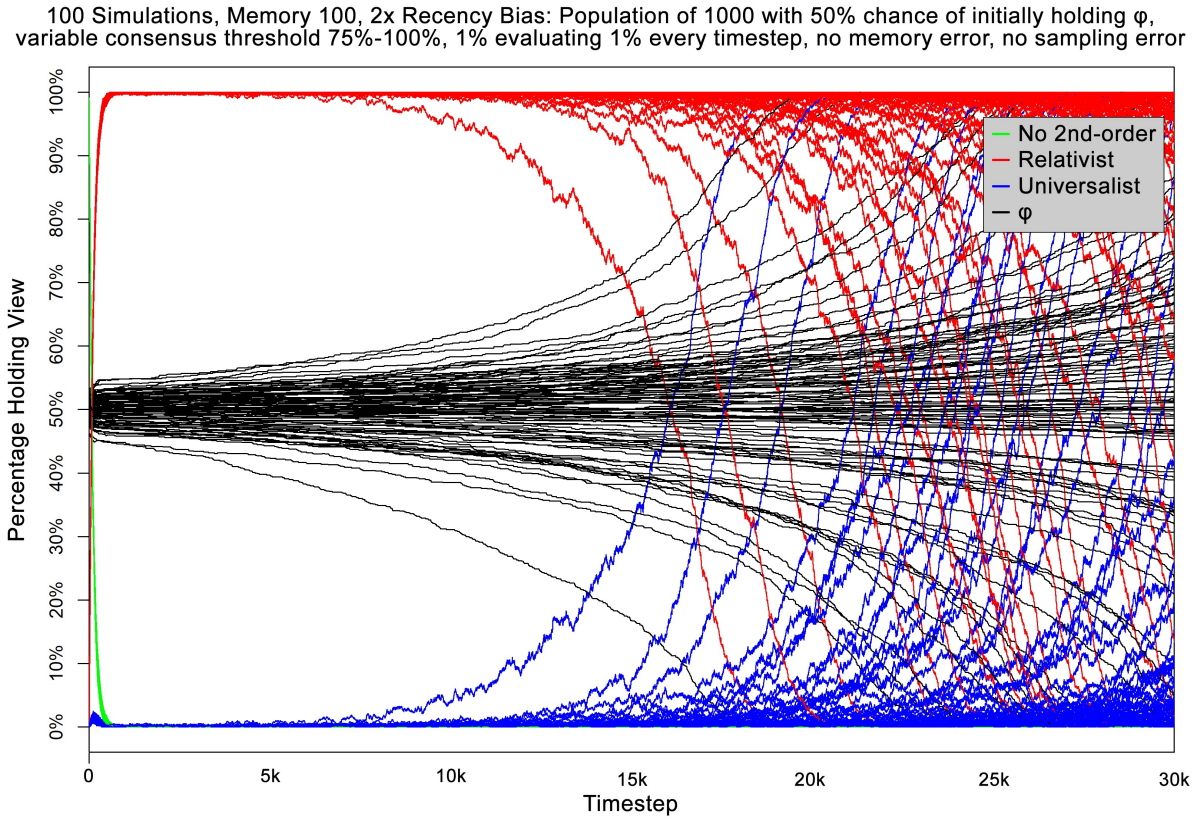


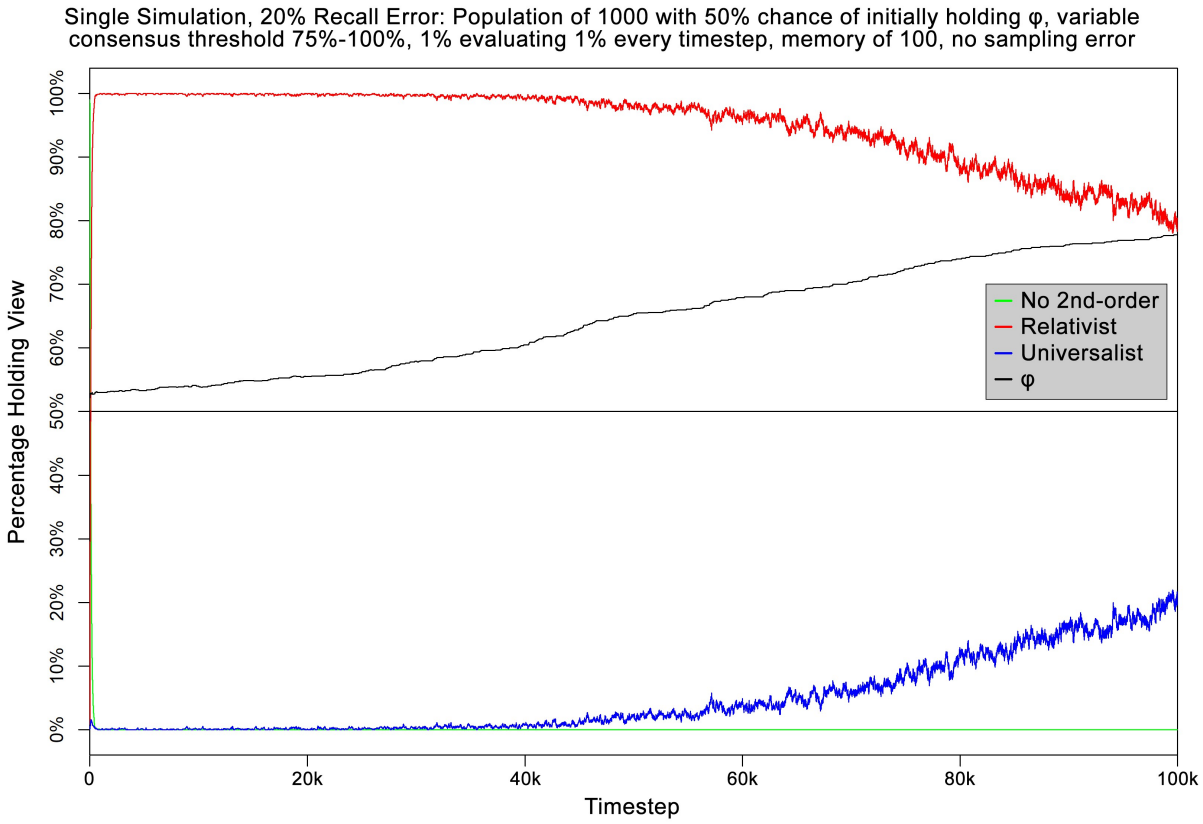
Figure 8: Simulations 8, Recency Bias

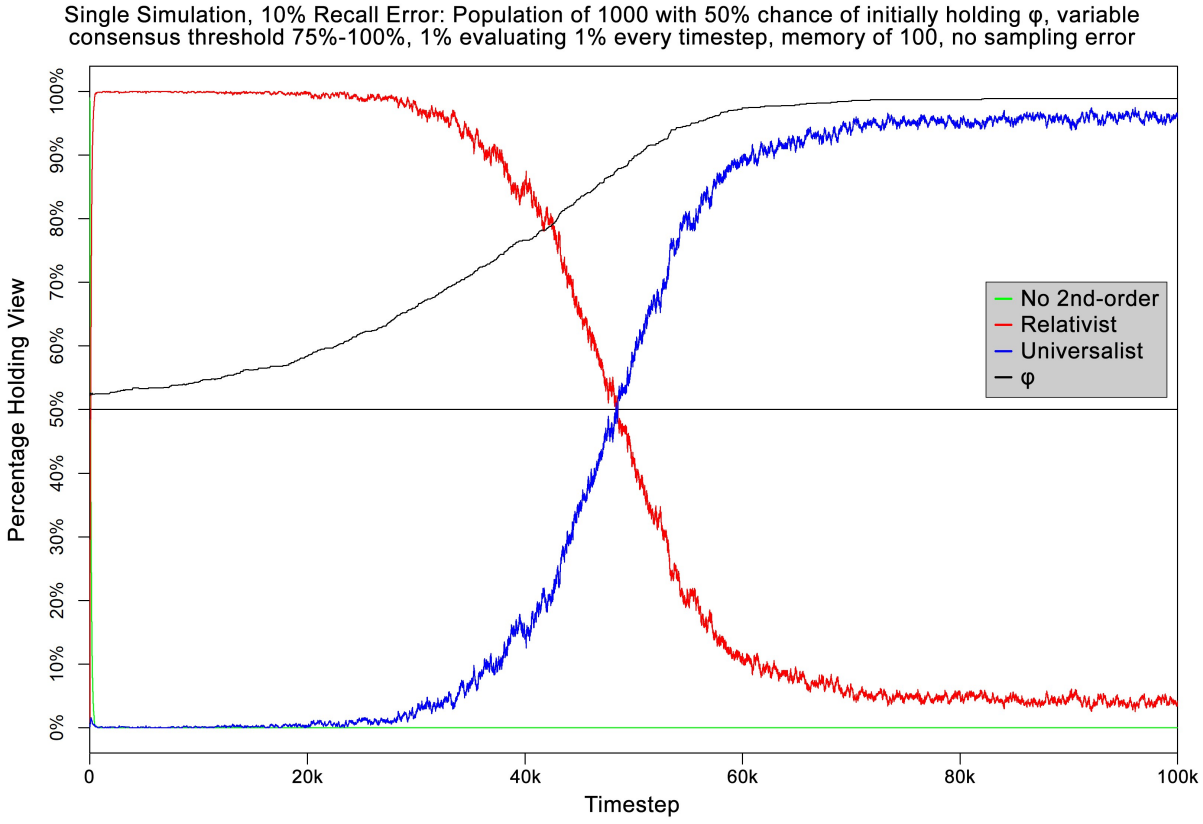
3.7 Memory Error

So far, we have varied how many past evaluations people call on in assessing consensus and how they weigh the evaluations that they call on. In these simulations, however, when people considered past evaluations they did so with perfect recall. This too is unrealistic—human memory is fallible. There are a number of further facets of memory that could be modeled. Most simply, we could include a recall error, with people having some probability of misremembering any given evaluation. To test this, we ran a set of single simulations using the same starting population as before, with a 2x recency bias and a memory of 100, adding in one of three levels of error—20%, 10%, and 2.5%. For each recalled evaluation, there was a given percent chance of changing from φ to not-φ or not-φ to φ. For purposes of comparison, each simulation was run

with 100k timesteps (Figure 9). We find either convergence on 100% φ, or a clear trend in that direction, for each error rate. We also find a move to universalist second-order beliefs in each case, although this is notably slower as the error rate increases.

Another issue these simulations draw out is that the consensus threshold people employ needs to take into account the possibility of error. If it doesn't, then with a sufficiently high threshold and error rate, people will be unlikely to come to a universalist second-order belief even if the population uniformly holds the same first-order view. This can be seen in the simulation with a recall error rate of 10%. Here we found that roughly 5% of the population at any given time still had relativist second-order beliefs even after first-order beliefs reached 100% φ. The reason here is clear: in a population where everyone holds φ but where there is 10% recall error, considering 100 evaluations (with no recency bias) will mean that on average nine errors will occur, and all errors will shift the assessed first-order belief from φ to not-φ. That means that if the average number of errors occurs, a person with a threshold higher than 91% would adopt relativism. And in these simulations we've been using a variable threshold ranging from 75% to 100%. As such, now that we've built error into the account, we'll shift the upper threshold down to 95% to help compensate for the possibility of error.


Single Simulation, 20% Recall Error: Population of 1000 with 50% chance of initially holding φ, variable consensus threshold 75%-100%, 1% evaluating 1% every timestep, memory of 100, no sampling error
Percentage Holding View
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0
20k
40k
60k
80k
100k
Timestep
No 2nd-order
Relativist
Universalist
φ


Single Simulation, 10% Recall Error: Population of 1000 with 50% chance of initially holding φ, variable consensus threshold 75%-100%, 1% evaluating 1% every timestep, memory of 100, no sampling error
Percentage Holding View
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0
20k
40k
60k
80k
100k
Timestep
No 2nd-order
Relativist
Universalist
φ

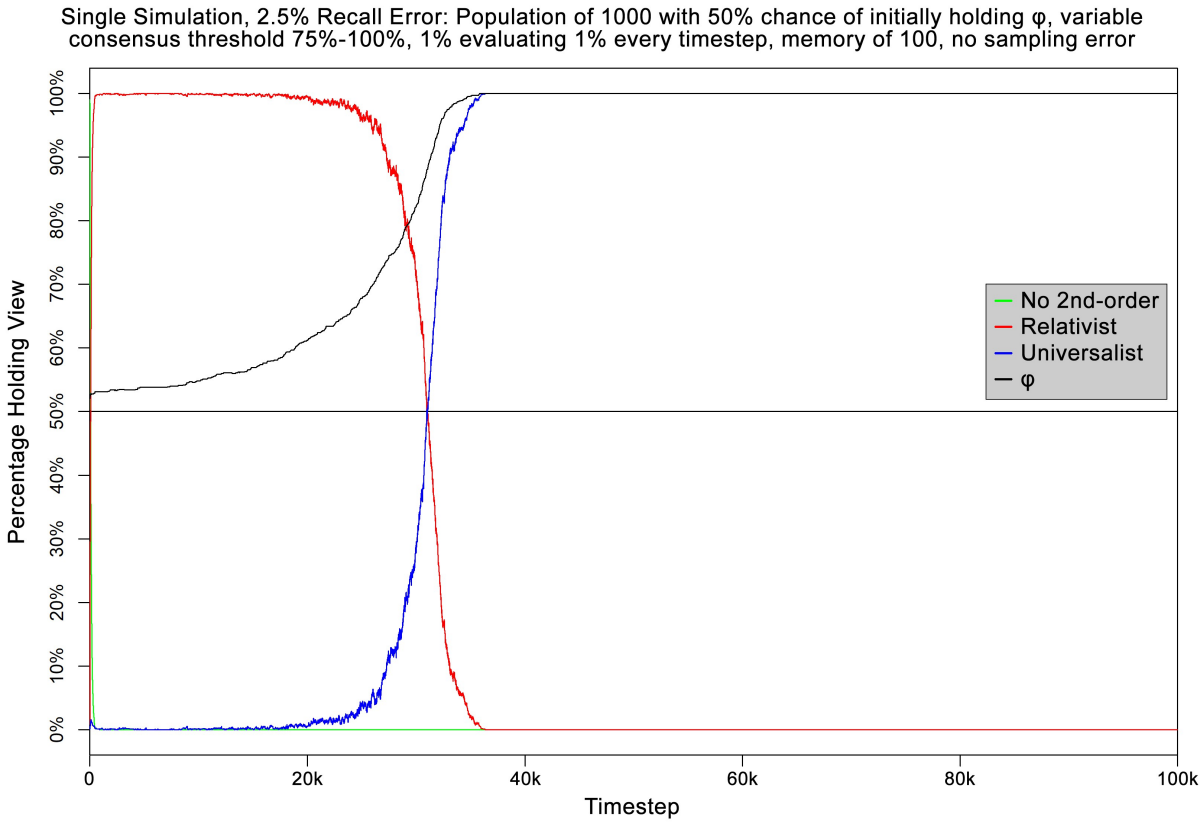


Figure 9: Simulations 9, Recall Error

While recall error serves to reduce the effect of second-order beliefs on first-order beliefs, it raises another issue for Nichols's account. As noted above, one of the substantive assumptions of his process vindication argument is that people must be relatively good at evaluating the truth-value of first-order moral claims. Although Nichols doesn't note this, a related assumption is that people must also be relatively good at discerning the results of other people's evaluations of first-order moral claims (i.e., their first-order beliefs). This holds for the same reason: what matters is the quality of the evidence that people are calling on in forming their second-order beliefs, and the quality of this evidence could be impacted by either errors in the source (if people make mistakes in assessing the truth-value of first-order moral claims) or in transmission (if people make mistakes in assessing what first-order beliefs other people hold). The latter can be thought of as sampling error and will be looked at below. Memory errors are distinct from

this, although they are potentially problematic for Nichols's account for the same basic reason: if people are calling on their memory of past evaluations of other people's first-order beliefs in assessing consensus, then memory errors would compromise the quality of that evidence.

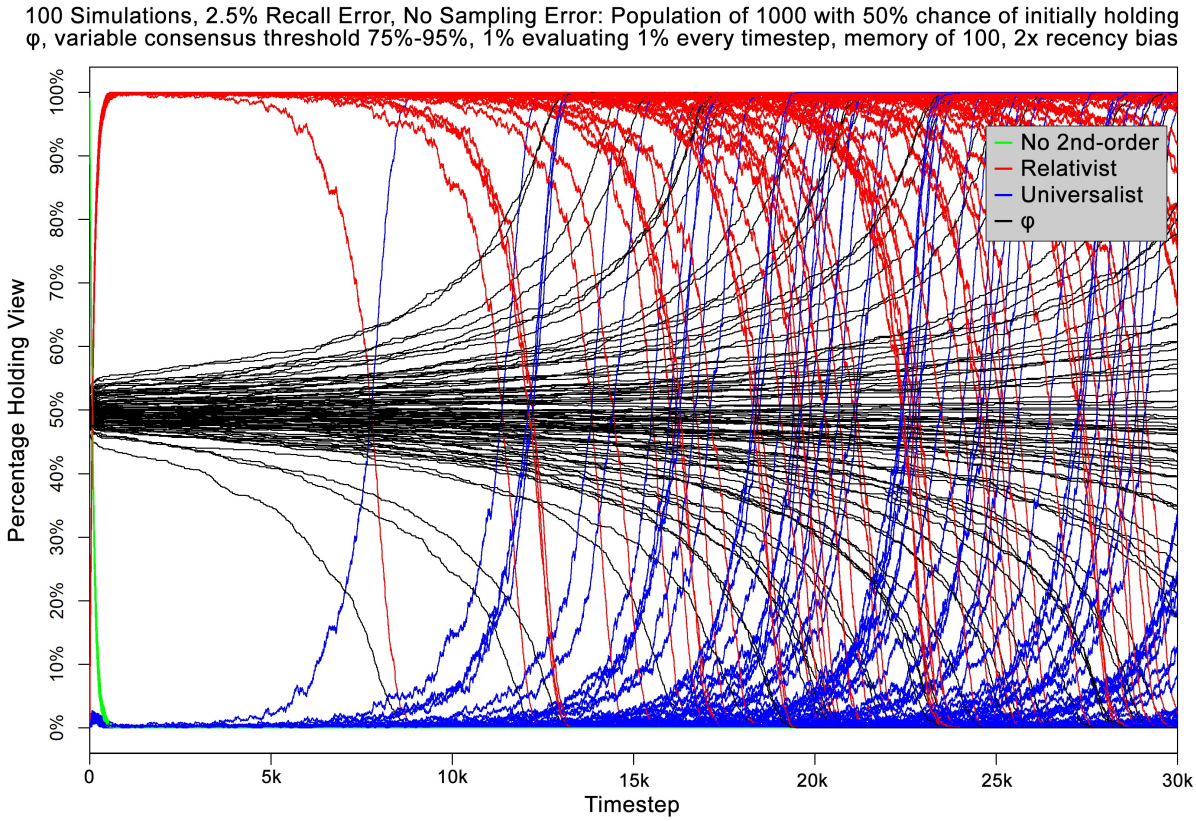
Nichols doesn't put any bounds on just how good people need to be at assessing first-order moral claims and determining those bounds is not a task we can take up here. Nonetheless, given the probability that there are also source and transmission errors involved, we think it is safe to assume that more than minimal recall error would be problematic for Nichols's account, at least for purposes of vindicating lay metaethical beliefs. Further, this issue would be magnified if people were prone to other types of memory errors. As such, we'll assume a relatively small recall memory error of 2.5% going forward.

3.8 Sampling Error

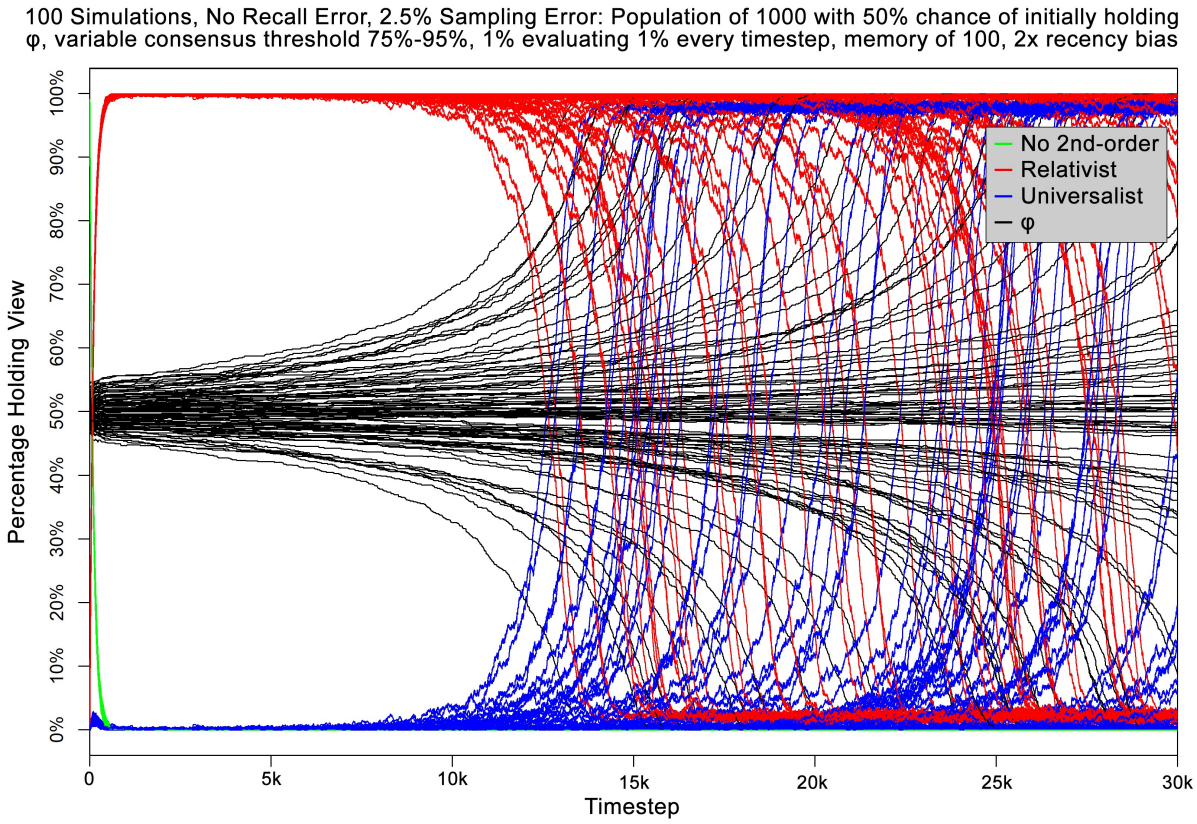
As just noted, just as it is likely that people will sometimes make errors in recalling past assessments, it is also likely that they will sometimes make errors in assessing the first-order beliefs of others. And as with memory errors there are different ways sampling error could work. As above, however, we'll set complications aside and model just a simple random sampling error where each new evaluation has a set chance of changing to the opposite belief from what the person being evaluated actually holds. Extensive sampling error would raise the same sorts of issues for Nichols's account as discussed for memory error, and as such we'll assume that sampling error is relatively modest, setting it the same 2.5% as recall memory error.

For purposes of comparison, we began by running 100 simulations using the parameters from the third simulation in Figure 9. We then ran another set of 100 simulations replacing the recall memory error with a 2.5% sampling error. Finally, we ran a set of 100 simulations with both a 2.5% recall memory error and a 2.5% sampling error. The results are shown in Figure 10.

What we find is that compared to simulations with no error, convergence is delayed when there is error, with there being a stronger effect for sampling error than recall error and the two together increasing the effect.[10] Our seventh important finding is that error increases the stability of relativism (although this comes at the expense of potentially compromising an appeal to the wisdom of the crowd).



[10] The greater effect of sampling error in our simulations presumably reflects that the sampling errors are persistent as they shift into memory and that the present sample is weighted more heavily when recency bias is present.



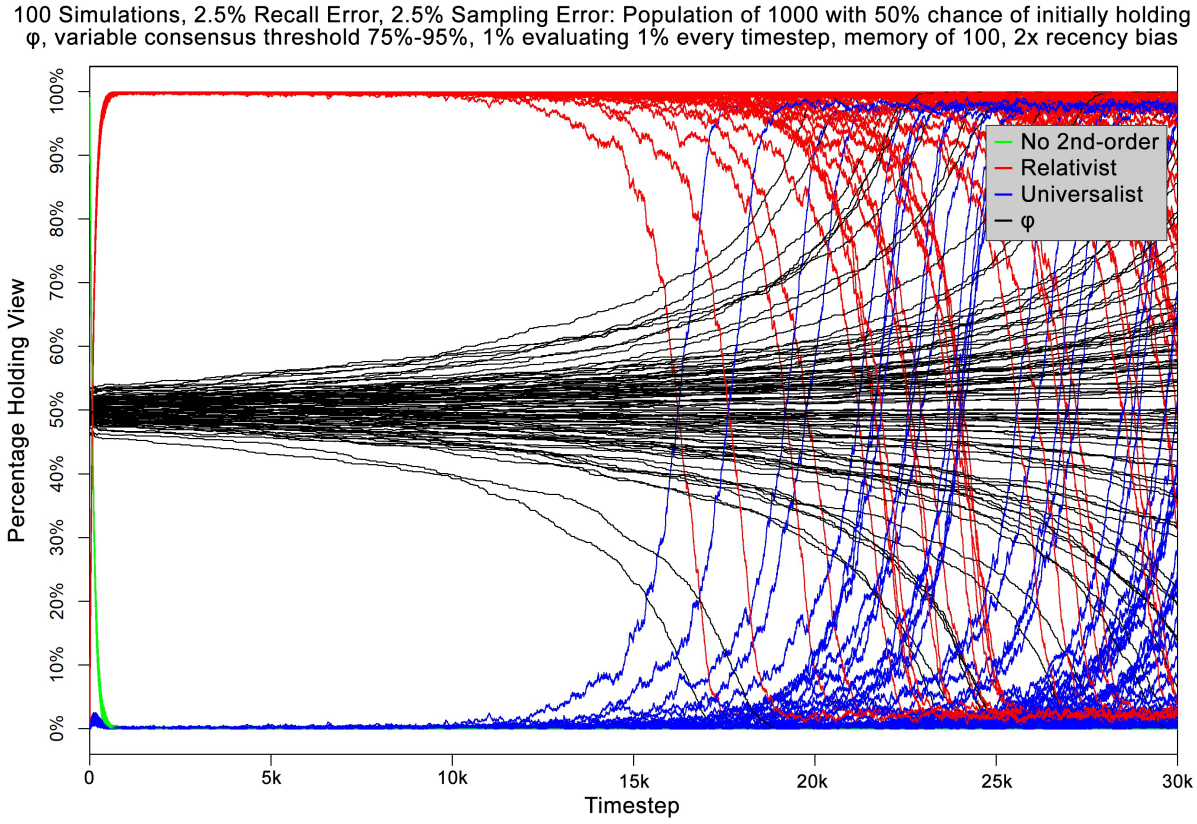


Figure 10: Simulations 10, Recall and Sampling Error

3.9 Local Sampling

So far, all of the simulations we've run have used what we might term *global* sampling: each person in the population randomly samples from the rest of the population in assessing first-order beliefs. It is more plausible that people would tend to sample *locally*, being more likely to assess the beliefs of people near them in the population. To test the effect of local sampling, we ran a single simulation with the parameters from the bottom set of simulations in Figure 10, but with each person sampling from just the 100 people closest to them in the population (Figure 11).[11] With this change we find that convergence isn't seen after even 100k timesteps.

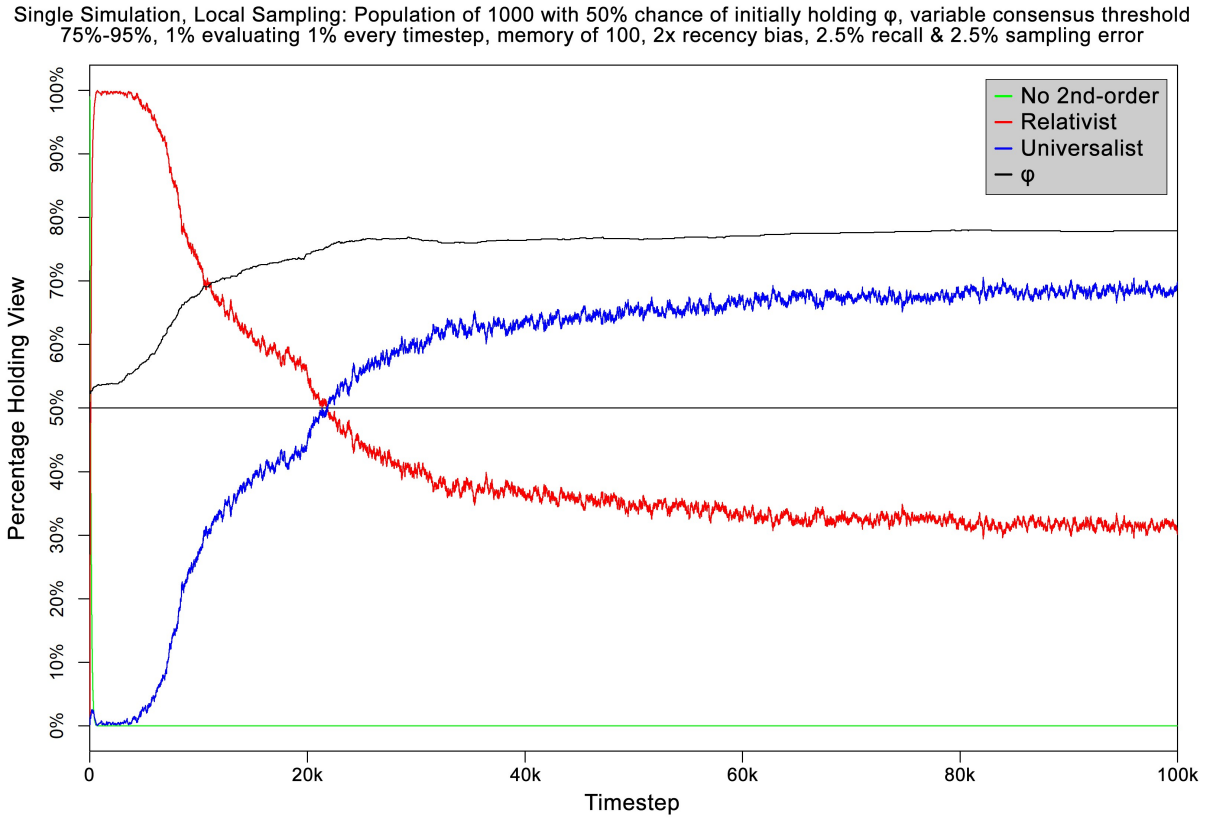
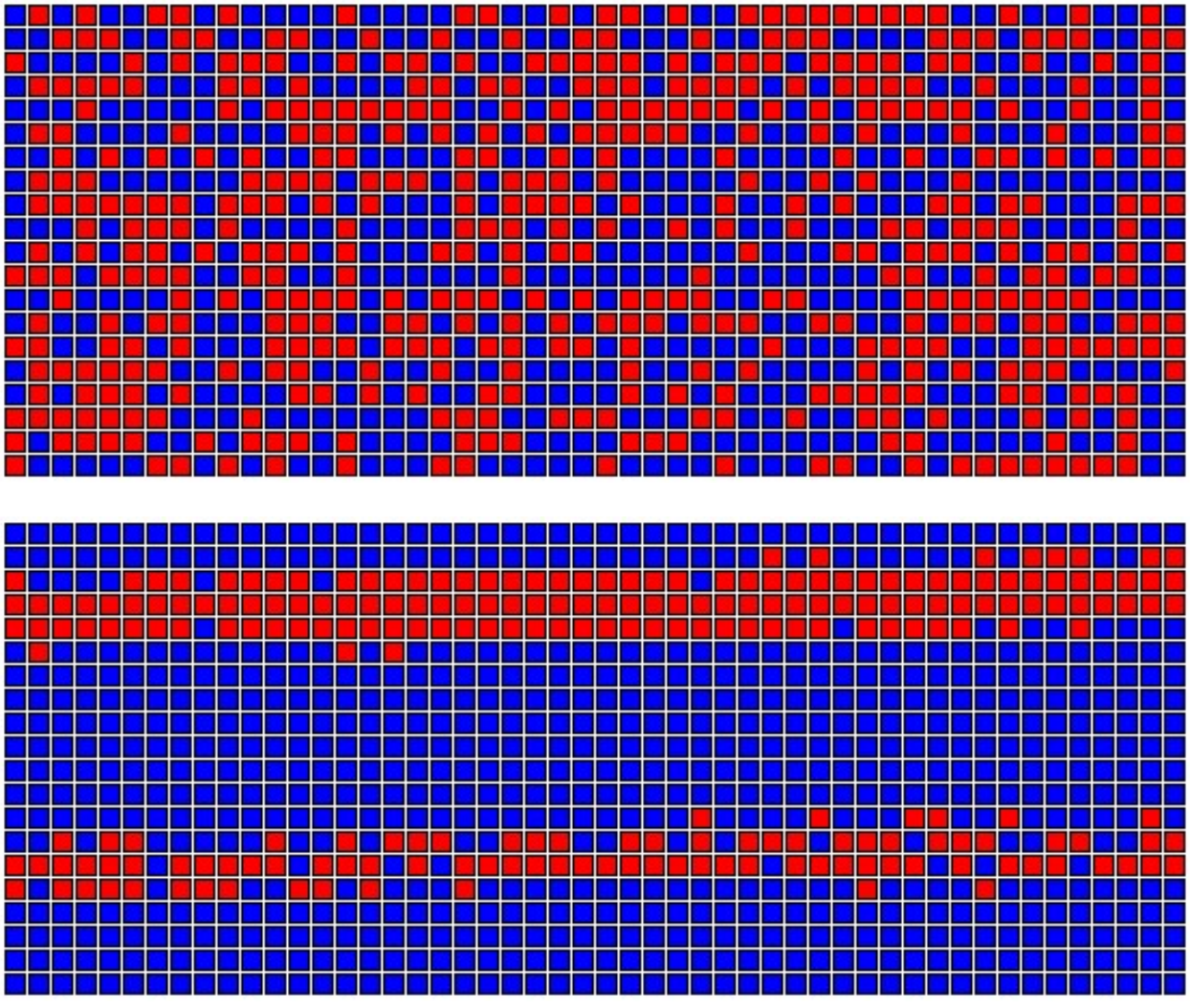


Figure 11: Simulation 11, Local Sampling

[11] For this we treated the population as a loop, such that the closest 100 people to the 1000th person in population are the first 50 people and the last 50 people excluding the evaluator.

The story here is more complicated than the previous description might suggest, however. Looking at how the population changes over time we find local bands of not-φ belief that are relatively stable over time. This is pictured in Figure 12, which shows the initial population followed by the population at time 50k and 100k. Our eighth important finding is that while local sampling can help maintain the presence of alternative first-order beliefs in the population, in our simulation it did so by generating pockets of dissent. Writ large this process would be likely to produce competing universalist beliefs, not to maintain a large number of relativists. An example of this is explored in 3.10.



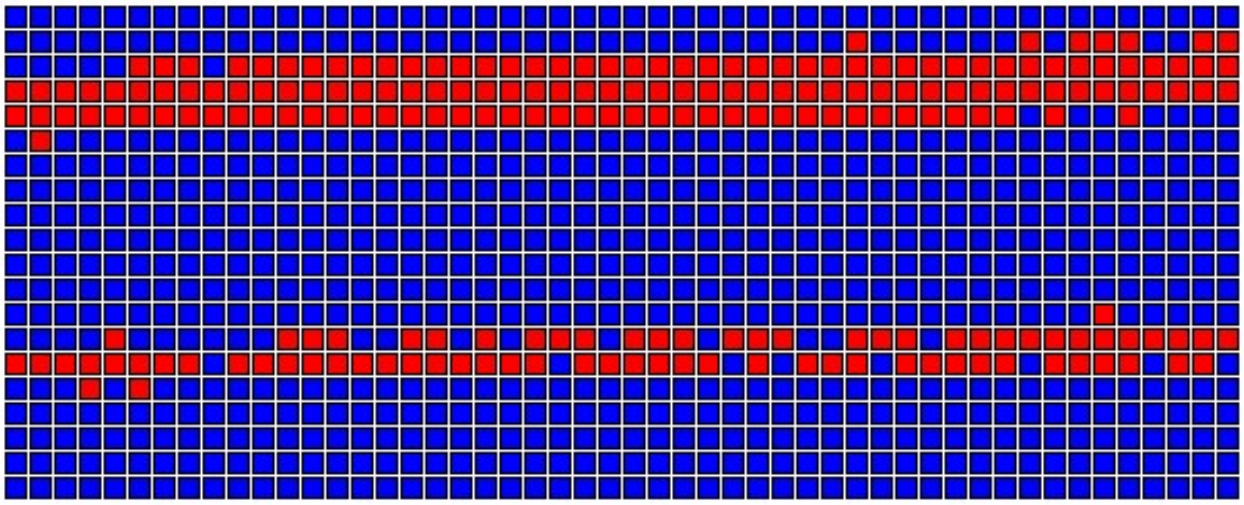
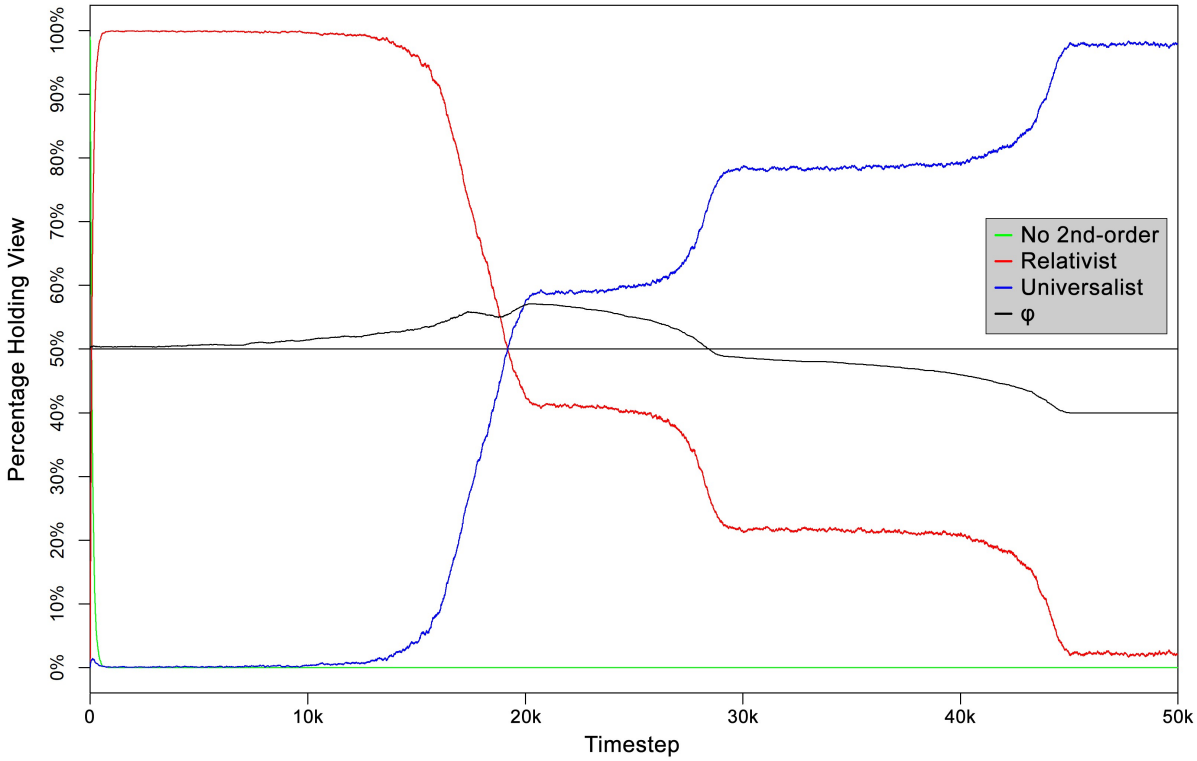


Figure 12: Visualization of the population from Simulation 11 at timestep 1 (top), 50k (middle), and 100k (bottom) with φ beliefs shown in blue and not-φ beliefs shown in red.

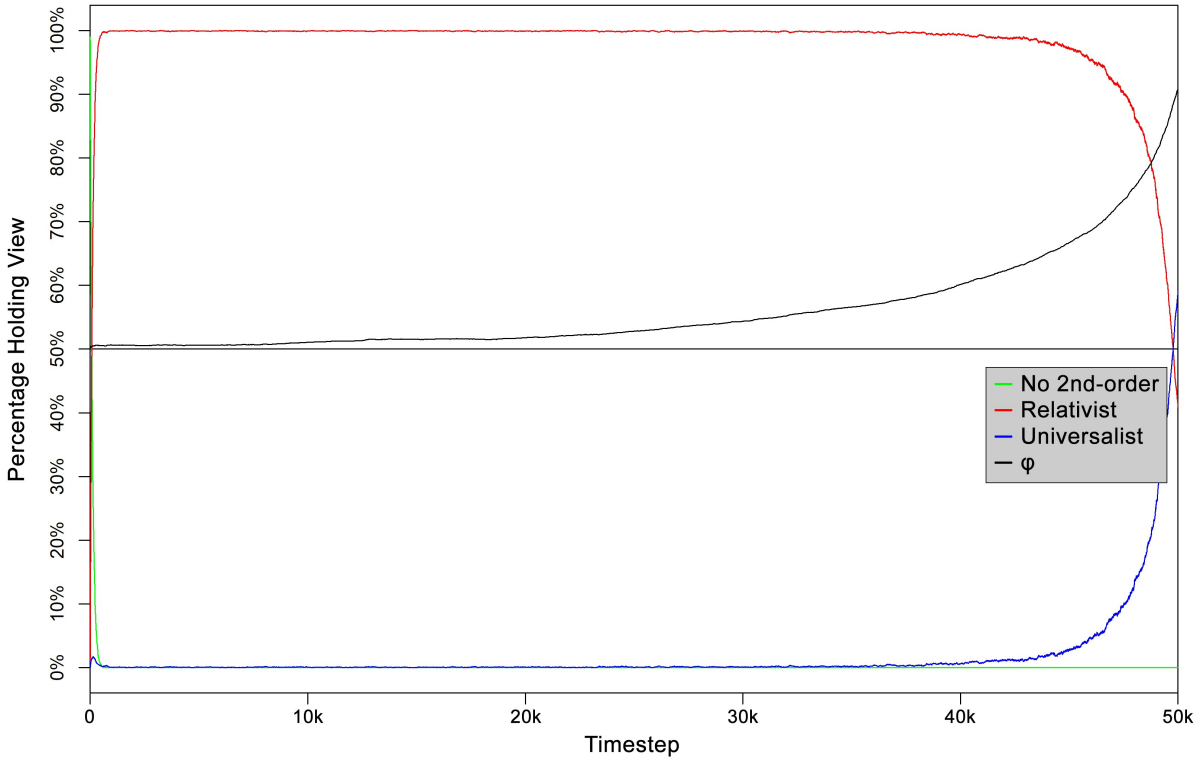
3.10 Multiple Populations and Migration

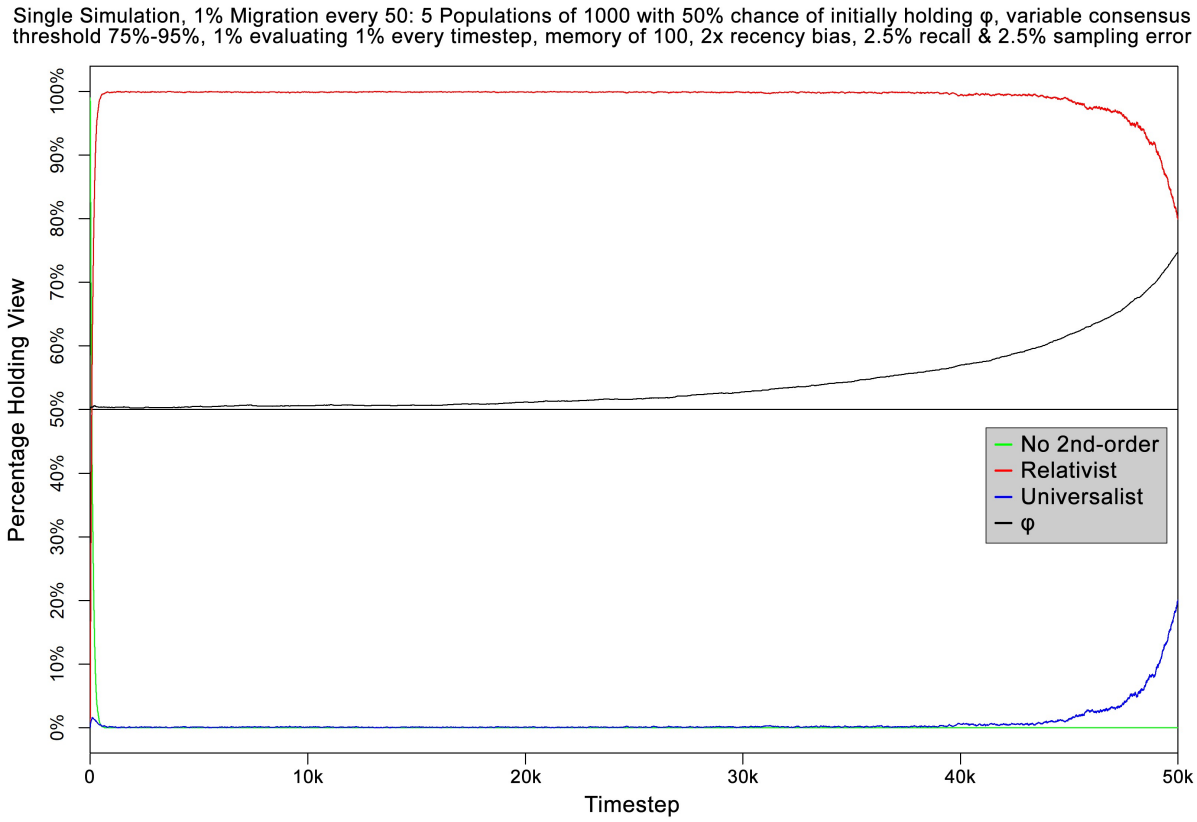
All of the simulations we've run so far have only modeled a single population at a time. More realistic would be to model multiple populations with the possibility of some flow of information between them. One way that information could flow is through migration, although this is certainly not the only one that occurs today (e.g., travel, telecommunications, the internet). Our final sets of simulations begin to look at the exchange of information between populations by modeling a simple form of migration. Our twelfth set of simulations model five populations of 1000 people where a random set of 1% of each population would periodically migrate to the next closest population in a loop. We then varied the frequency of migration, running a single simulation with the same starting populations for no migration, migration every 10 timesteps, migration every 50 timesteps, and migration every 100 timesteps. Each simulation was run over 50k timesteps and otherwise used the previous parameters (Figure 13).

Single Simulation, No Migration: 5 Populations of 1000 with 50% chance of initially holding φ, variable consensus threshold 75%-95%, 1% evaluating 1% every timestep, memory of 100, 2x recency bias, 2.5% recall & 2.5% sampling error



Single Simulation, 1% Migration every 10: 5 Populations of 1000 with 50% chance of initially holding φ, variable consensus threshold 75%-95%, 1% evaluating 1% every timestep, memory of 100, 2x recency bias, 2.5% recall & 2.5% sampling error





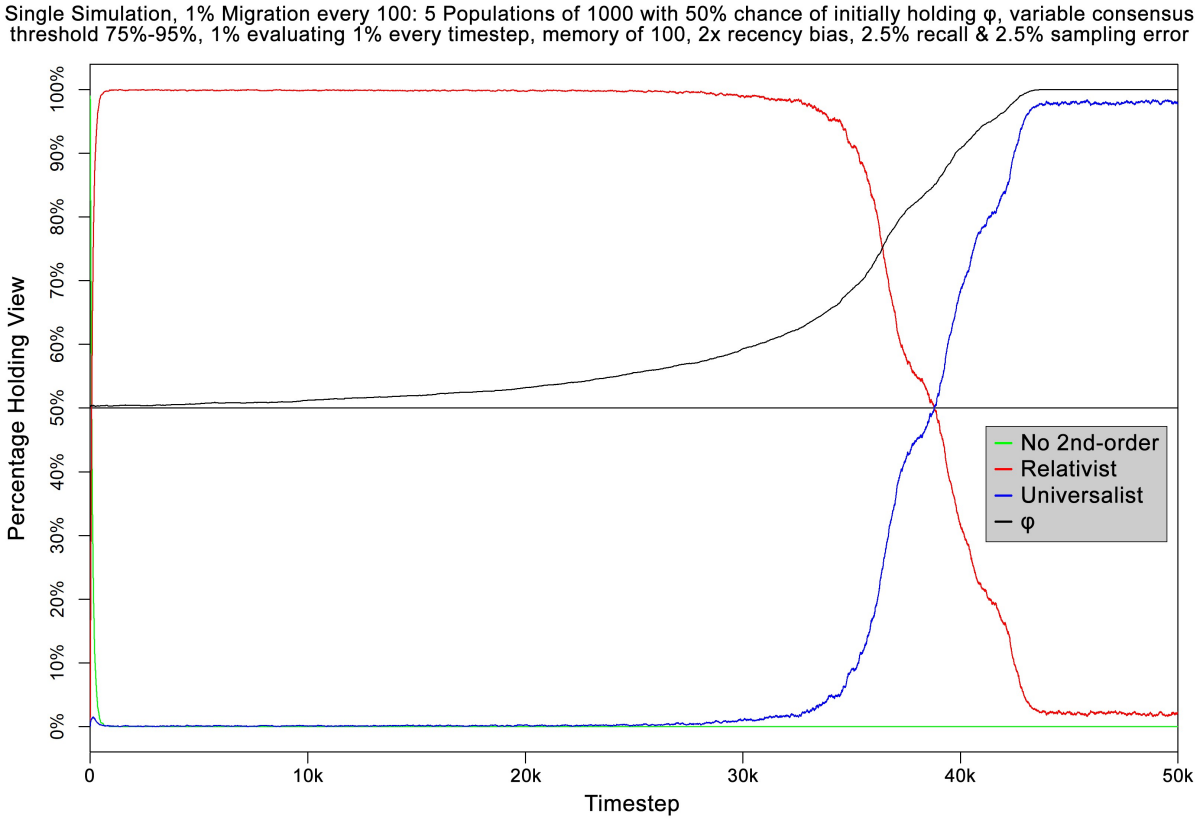
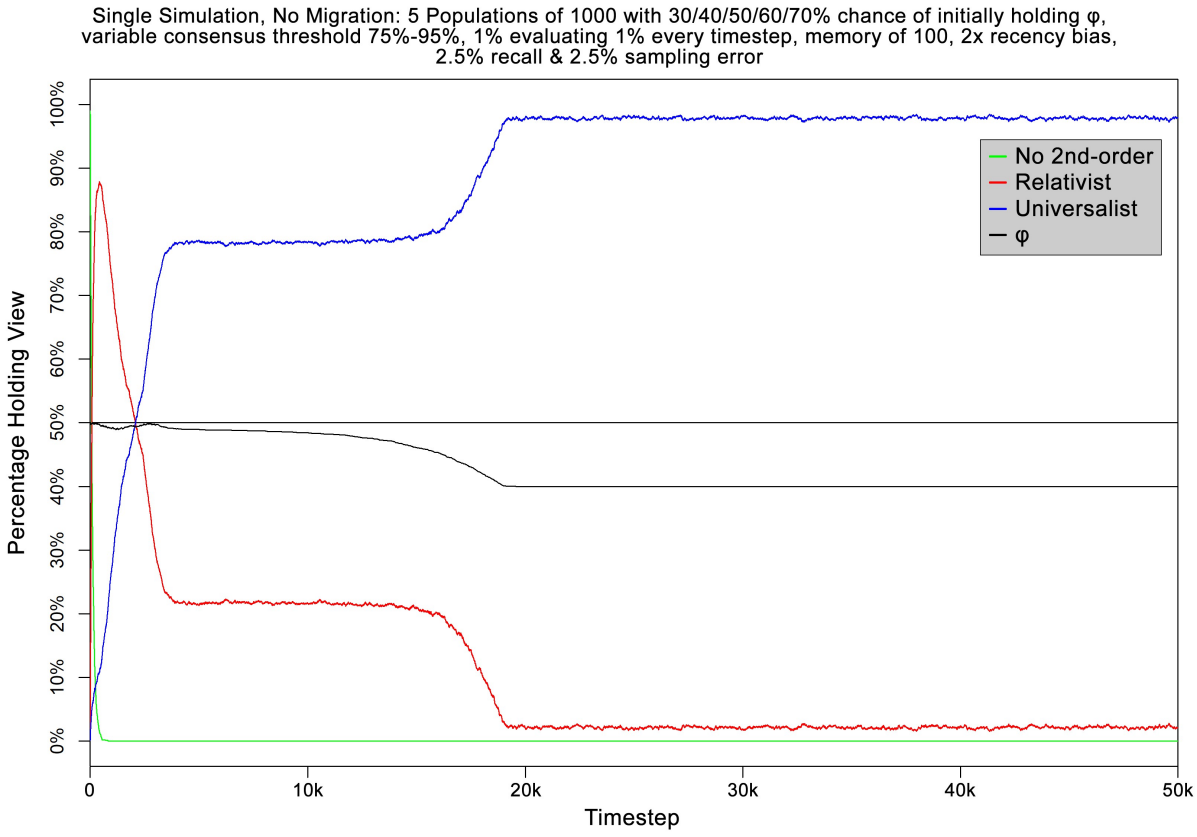
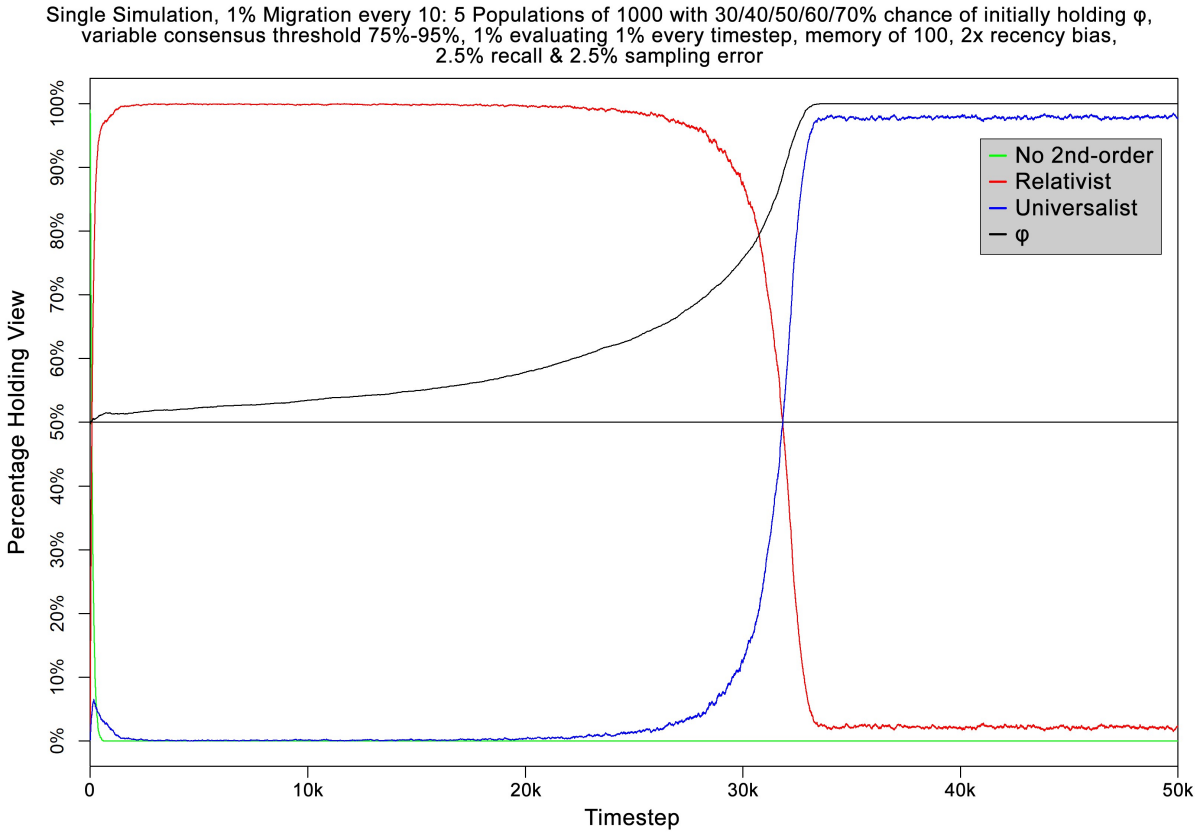


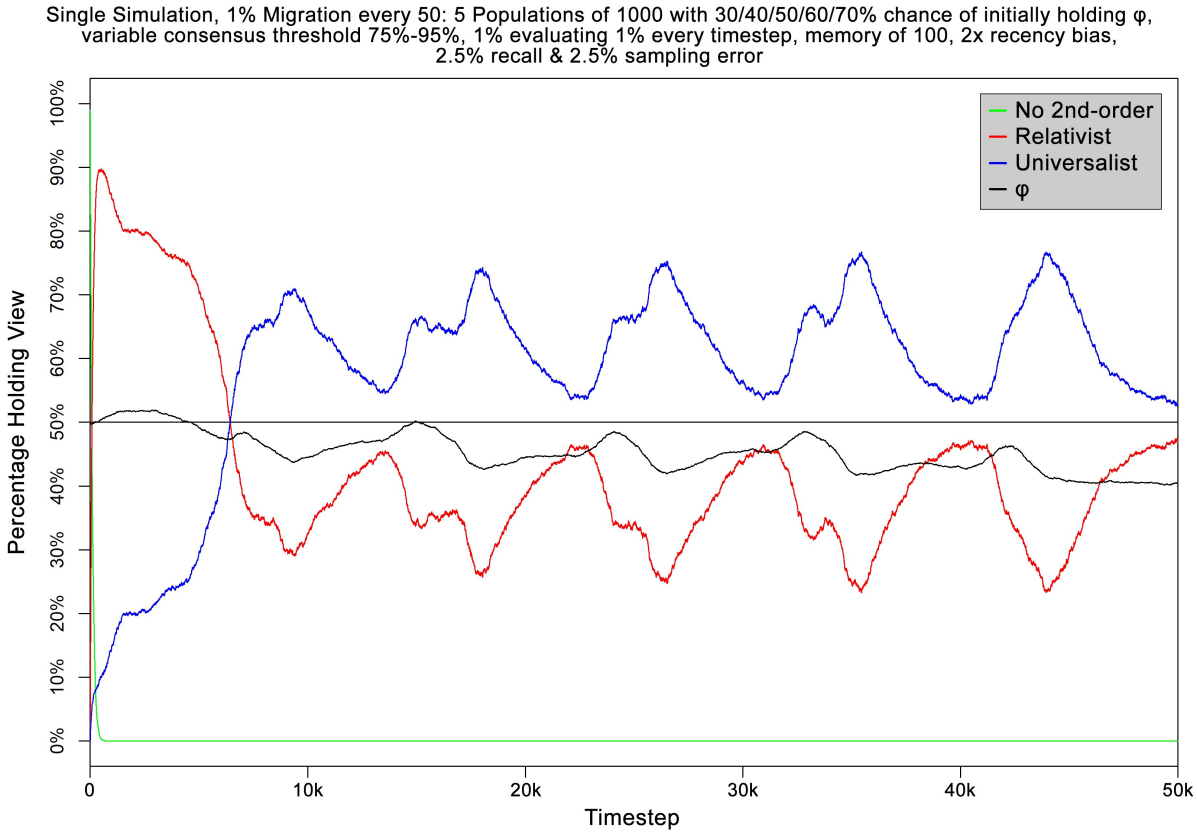
Figure 13: Simulations 12, Migration

Starting with the simulation with no migration, while we do not find convergence to one first-order belief for all of the populations taken together, each individual population converges on either 100% φ (Populations 1 and 2) or 100% not-φ (Populations 3-5), and as a result universalist beliefs approach 100%. Further, the sharp rises in the plot separated by plateaus indicate that the populations do not all converge at the same rate, with there being a slow increase in the dominant belief before it hits a point where first-order beliefs cascade. Adding in migration, we see that this delays convergence, with only the least frequent pattern of migration (1% every 100 timesteps) showing convergence within 50k timesteps. For that simulation, we found that all five populations converged on 100% φ. The implication is that the delay is due to what we might think of as warring belief systems: without migration three of the populations converged on not-φ, but with infrequent migration those populations were instead pulled to φ. A similar pattern is evident for the other two simulations with less frequent migration, although convergence is delayed. For the simulation with frequent migration (1% every 10 timesteps), after 50k timesteps all five populations were coming close to 100% φ (Population 1: 95.3%; Population 2: 89.0%; Population 3: 85.9%; Population 4: 88.5%; Population 5: 95.8%). For the simulation with a medium migration frequency (1% every 50 timesteps), all five populations were moving toward 100% φ, although the percentage of φ beliefs was lower in four of the five populations than it was when there was more frequent migration (Population 1: 66.2%; Population 2: 61.4%; Population 3: 77.2%; Population 4: 91.4%; Population 5: 78.4%).

To further investigate the occurrence of "warring belief systems," in our thirteenth set of simulations we added "cultural variation" in first-order beliefs to the simulations, varying the likelihood that people from each population would initially hold φ. We began by looking at a distribution of likelihoods of 70% for Population 1 through 30% for Population 5, decreasing in 10 percentage point increments (Figure 14).


Single Simulation, No Migration: 5 Populations of 1000 with 30/40/50/60/70% chance of initially holding φ,
variable consensus threshold 75%-95%, 1% evaluating 1% every timestep, memory of 100, 2x recency bias,
2.5% recall & 2.5% sampling error
Percentage Holding View
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0
10k
20k
30k
40k
50k
Timestep
No 2nd-order
Relativist
Universalist
φ


Single Simulation, 1% Migration every 10: 5 Populations of 1000 with 30/40/50/60/70% chance of initially holding φ,
variable consensus threshold 75%-95%, 1% evaluating 1% every timestep, memory of 100, 2x recency bias,
2.5% recall & 2.5% sampling error
Percentage Holding View
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0
10k
20k
30k
40k
50k
Timestep
No 2nd-order
Relativist
Universalist
φ



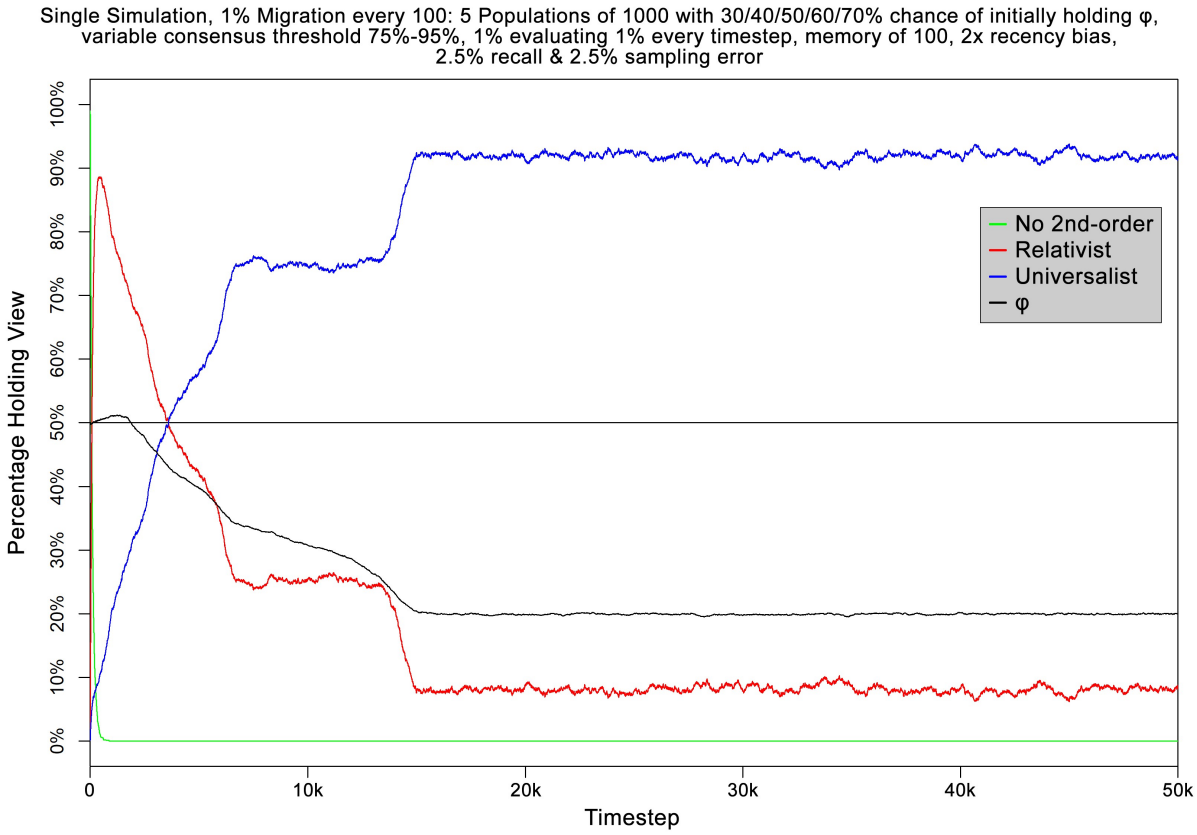


Figure 14: Simulations 13, Cultural Variation with Migration

Not surprisingly, we again find that convergence does not occur globally when there is no migration. Looking at the individual populations, we find that at 10k timesteps Populations 1 and 2 have converged to 100% φ, while Populations 4 and 5 have converged to 100% not-φ, and Population 3 is in the process of converging to not-φ (578/100 holding not-φ). After roughly 20k timesteps, however, all five populations have converged on one of the two first-order beliefs.

When migration is included, the pattern gets more complicated. With relatively frequent migration (1% every 10 timesteps), we found that the populations all converged to 100% φ. Looking at the populations after 10k timesteps, we find that they had all moved to slightly greater than 50% φ (Population 1: 53.6%; Population 2: 52.2%; Population 3: 52.3%; Population 4: 52.9%; Population 5: 56.5%). When migration is decreased to 1% every 50 timesteps, however, we find oscillation between a relatively high percentage of universalist beliefs (around 70%) and a more even split between universalist and relativist beliefs, although the percentage of φ beliefs appears to be slowly decreasing. Looking at the populations we see radically different percentages of φ beliefs both at 10k timesteps (Population 1: 85.8%; Population 2: 0.1%; Population 3: 2.5%; Population 4: 34.8%; Population 5: 99.4%) and at 50k timesteps (Population 1: 95.2%; Population 2: 35.5%; Population 3: 0.3%; Population 4: 6.0%; Population 5: 64.9%). We might think of this as an entrenched war of beliefs, with the minority belief only very slowly giving ground. Further, the occurrence of oscillation is robust across starting populations, as seen in Figure 15 which plots 100 simulations with the same parameters.

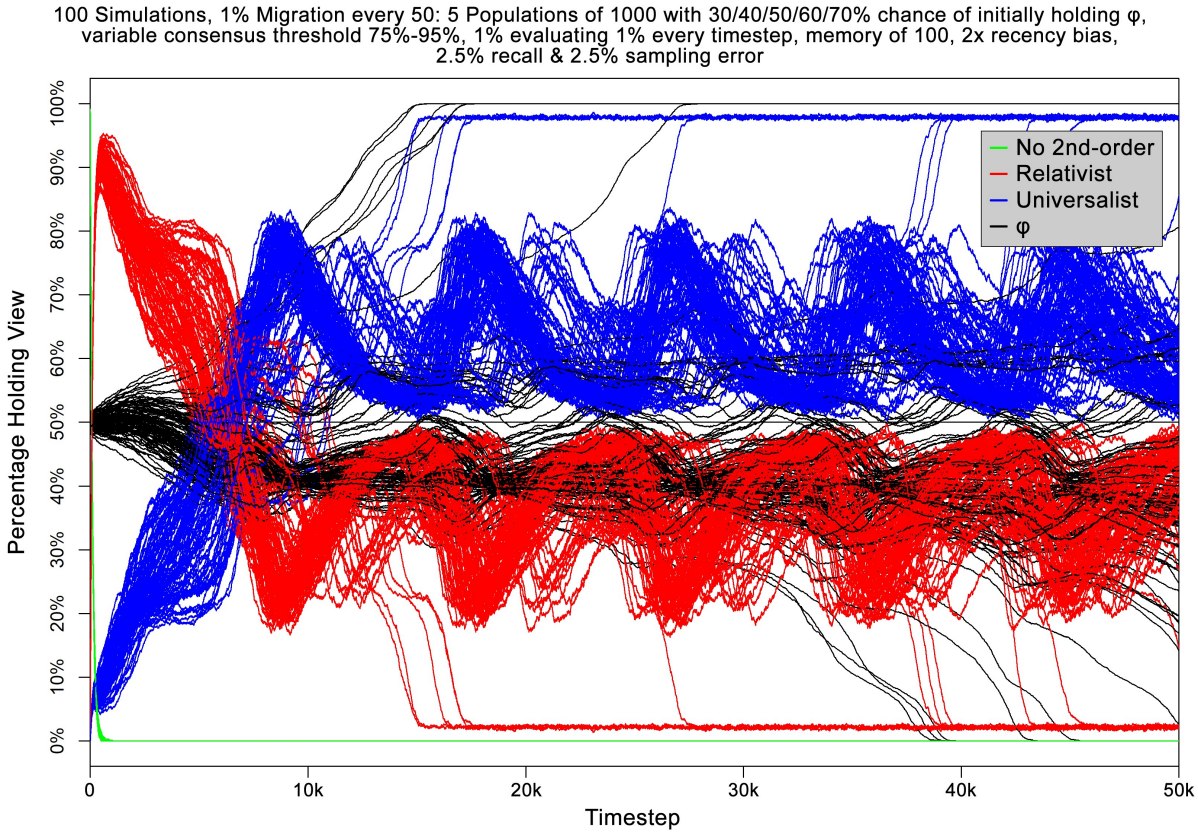


Figure 15: Simulations 14, "Warring Belief Systems"

Finally, when we decrease the frequency of migration still further to 1% every 100 timesteps, we find a quick rise in universalist beliefs, followed by a plateau around 75% universalist beliefs, followed by another quick rise and an extended plateau at just over 90% universalist beliefs. At that point φ beliefs stay relatively steady at just over 20%. Looking at the populations, at 10k timesteps—in the middle of the first plateau—we find a divide between the populations, with Population 1 nearly all holding φ (97.6%), Population 2 close to evenly split (51.0%), and the remaining populations nearly all holding not-φ (Population 3: 0.0%; Population 4: 0.1%; Population 5: 5.0%). At 50k timesteps—in the middle of the second plateau—things are similar, although Population 2 has now converged on 0% φ beliefs (Population 1: 94.5%; Population 2: 0.0%; Population 3: 0.0%; Population 4: 0.1%; Population 5: 5.2%).

The ninth and final important finding we flag is that information flow between populations, represented here by simple directional migration, has a complex effect on first-order beliefs. Since first-order beliefs in separate populations tend to randomly converge on one or the other of the two beliefs, without migration multiple populations are likely to be moving in different directions. Migration then puts these trends into tension, delaying convergence.

4. Conclusion

In this paper we looked at Nichols's recent process vindication argument for lay metaethical beliefs. This argument proposes that people are sensitive to consensus information in forming metaethical beliefs and contends that calling on consensus information in this way is at least locally computationally rational. This vindication of lay metaethical beliefs hinges on a number of substantive assumptions, however, including that people's first-order beliefs are to a large extent independent. We raised two concerns with regard to this assumption. First, empirical work suggests that people call on their assessment of the first-order moral beliefs of others in forming their own first-order beliefs. Second, if people do call on consensus information in forming second-order beliefs, and if they maintain consistency between their first-order beliefs and their second-order beliefs, then independence will be threatened by the use of the very process at issue. To begin to investigate whether the latter is a live worry for Nichols's process vindication argument, we ran a series of computer simulations to explore the circumstances under which the use of consensus information in this way is likely to cause a problematic failure of independence such that belief in relativism becomes unstable over time despite an initial lack of consensus with regard to a first-order moral claim.

While our simulations use highly simplified models, and while the best parameters for these models will depend on what empirical testing reveals about actual human belief formation,

the findings nonetheless suggest that independence cannot be safely assumed with regard to the process Nichols discusses. This in turn raises doubts about whether this process is rational, and with it whether the occurrence of such a process would serve to vindicate lay metaethical beliefs.

Moving beyond vindicating lay metaethical beliefs, we might wonder why people form second-order beliefs that are sensitive to perceived consensus information, as the empirical evidence suggests that they do. Our simulations suggest a potential answer: forming second-order beliefs in this way promotes convergence in first-order beliefs, which will be beneficial when coordination is desirable. Expanding on this, we would predict that a tendency toward universalism for a given issue or domain would tend to speed up convergence in first-order beliefs, while a tendency toward relativism would tend to slow it down, maintaining diversity. This possible role for lay metaethical beliefs will be explored in subsequent work.

References

Ayars, A., & Nichols, S. (2019). Rational learners and metaethics: Universalism, relativism, and evidence from consensus. *Mind & Language*, https://doi.org/10.1111/mila.12232

Cameron, C., Payne, B., & Doris, J. (2013). Morality in high definition: emotion differentiation calibrates the influence of incidental disgust on moral judgments. *Journal of experimental social psychology, 49*(4), 719-725.

Dietrich, F., & Spiekermann, K. (2013). Independent Opinions? On the Causal Foundations of Belief Formation and Jury Theorems. *Mind, 122*(487), 655-685.

Goodwin, G. P., & Darley, J. M. (2008). The Psychology of Meta-Ethics: Exploring Objectivism. *Cognition, 106*(3), 1339-1366.

Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs percieved to be more objective than others? *Journal of Experimental Social Psychology, 48*(1), 250-256.

Joyce, R. (2001). *The Myth of Morality.* Cambridge: Cambridge University Press.

Joyce, R. (2011). The Error In 'The Error In The Error Theory'. Australasian Journal of Philosophy, 89(3), 519-534.

Marr, D. (1982). *Vision.* MIT Press.

Nichols, S. (2019). Debunking and Vindicating in Moral Psychology. In A. I. Goldman, & B. P. McLaughlin (Eds.), *Metaphysics and Cognitive Science* (pp. 99-122). Oxford University Press.

Nichols, S. (2019). Experimental Philosophy and Statistical Learning. In E. F. Curtis (Ed.), *Methodological Advances in Experimental Philosophy* (pp. 13-41). London: Bloomsbury Academic.

Nichols, S., Roojen, M. v., & Murray, D. (forthcoming). Minimal objectivism and evaluative properties.

Stein, E. (1996). *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science.* Oxford: Oxford University Press.

Sytsma, J., & Livengood, J. (2016). *the theory and practice of Experimental Philosophy.* Ontario: Broadview Press.

Wright, J. C., Grandjean, P. T., & McWhite, C. B. (2013). The meta-ethical grouding of our moral beliefs: Evidence for meta-ethical pluralism. *Philosophical Psychology, 26*(3), 336-361.

Ziegelmeyer, A., Koessler, F., Bracht, J., & Winter, E. (2010). Fragility of information cascades: an experimental study using elicited beliefs. *Experimental Economics, 13*(2), 121-145.