Reliability of Testimonial Norms in Scientific Communities

Conor Mayo-Wilson

April 11, 2013

Introduction

Most of our scientific knowledge is based upon the testimony of others. For example, I have never verified Coulomb's law in a laboratory, but I still know that electrostatic force obeys an inverse square law. Why? My high school physics teacher told me so. On first glance, explaining my knowledge of such scientific facts seems rather easy. Scientific facts, like Coulomb's law, are confirmed repeatedly by experts in laboratories or other experimental settings. Experts then disseminate their findings by word of mouth or through journal articles. Those findings are, in turn, summarized in survey articles for other academics, in textbooks for high school students, and in popular articles and lectures for the lay public. In this way, experts' knowledge of scientific facts, like Coulomb's law, is transmitted from one person to another. Each part of this explanation, however, raises serious epistemological questions. I will mention three.

First, scientists often disagree. In the presence of such disagreements, how can non-specialists justifiably accept a scientific hypothesis merely on the basis of others' claims, especially when there are experts who hold an opposing view? Recognizing the importance of resolving conflicting expert testimony, epistemologists and legal theorists have begun to develop procedures for evaluating experts and deciding who to trust. Some epistemologists have argued that, absent other information, non-specialists should "go by the numbers" and adopt the opinion of the majority of experts in a field. Others have argued that there are a series of criteria by which non-experts can determine which experts are most reliable.[1]

[1] The Lehrer-Wagner model entails that, all other things being equal, greater weight ought to be assigned to beliefs that are held by many experts rather than a few. See Lehrer

Second, much of our scientific knowledge is acquired from non-experts.[2] Journalists often have no scientific training, and yet both academics and lay audiences often learn of scientific advances via newspapers and magazines. Many secondary teachers and college professors are not experts in the field in which they provide instruction. And so on. Given that non-experts are sometimes unreliable, may have strong incentives for dishonesty, and are prone to miscommunication, there are even more philosophical and practical questions concerning who one can justifiably trust.[3] Recent debates in the epistemology of testimony have led to the establishment of two positions, called reductionism and non-reductionism respectively, that attempt to characterize the conditions under which one is justified in accepting others' claims. Recognizing the implausibility (or our inability) of always verifying others' claims, non-reductionists argue that one can justifiably trust a speaker in the absence of evidence of dishonesty or unreliability. In contrast, recognizing the frequency of dishonest and/or unreliable communication, reductionists claim that one needs *positive reasons* to trust others, where such positive reasons might include evidence for the speakers' honesty and/or expertise an area. And there are philosophers who adopt intermediate positions.[4]

Finally, several philosophers have questioned whether *knowledge* is trans-

and Wagner [1981]. In contrast, Goldman [2001] argues that, because experts judgments might be highly correlated due to common information, agreement cannot always provide greater evidence of a hypothesis. As a result, Goldman claims that there any number of heuristics that one might use to evaluate expert testimony.

[2] I am not claiming there is a bright line between experts and non-experts; there are clearly degrees of expertise and knowledge. My point is that we often rely on individuals who are not extremely well-versed in the area in which they are testifying. Moreover, I would hypothesize that we often rely on individuals who are no more qualified to evaluate the truth of scientific hypotheses than ourselves.

[3] I am not claiming that scientists are immune from error, dishonesty, and/or miscommunication. There are plenty of cases to indicate otherwise. Rather, I am claiming that universities, academic journals, and research institutions have established mechanisms to mitigate these sources of error, whereas such mechanisms are often not present in everyday conversational contexts, television journalism, and so on. Peer review, for instance, arguably decreases the chances of mistakes; there are strong punishments for fabricating data that minimize dishonesty; finally, the training of scientific professionals, along with the establishment of technical vocabularies, can decrease the chances of miscommunication. Such incentives and punishments are, I assume, less frequent outside of academic settings.

[4] See Burge [1993], Coady [1973], and Foley [2005] for several different defenses of the non-reductionist position. See Adler [1994], Fricker [1994], Fricker and Cooper [1987] for defenses of reductionism. An extensive list of citations for this debate is available in Lackey [2011].

ferred via testimony. Lackey [1999], for example, argues that a creationist biology teacher might not know the theory of evolution (because she does not believe it to be true) and yet still successfully teach it to students (who thereby acquire knowledge). Lackey's examples raise the possibility that we can know scientific facts via testimony even if those facts are transmitted via a series of conversational or written exchanges in which no individual believes, or has sufficient reason to believe, the fact in question.

Many of the above philosophical debates are (at least implicitly) motivated by concerns about the reliability of various rules for changing one's beliefs in light of others' claims. Call such rules **testimonial norms**. The rule "believe others' claims in the absence of conflicting information" is one testimonial norm, and the rule "believe all and only those individuals you know to be reliable and trustworthy" is another. Arguably, the debate between reductionists and non-reductionists is, in part, motivated by the observation that the former norm is reliable in certain contexts but not in others, and the latter norm prohibits one from learning from individuals in certain contexts in which speakers are generally reliable and trustworthy. Similar remarks apply to debates about expert testimony.[5]

It is surprising, therefore, that epistemologists have made little effort (i) to characterize those contextual features that influence the reliability of different testimonial norms, or (ii) to evaluate the reliability of different testimonial norms as those contextual features vary.[6] Both projects

[5] Testimonial norms are a subset of what are generally called *epistemic norms* (or *epistemic rules*). See Pollock [1987]. Two notes are important. First, most epistemologists have discussed only *passive* norms, i.e., rules for updating one's beliefs in light of evidence. However, there might be epistemic norms dictating *active* obligations to acquire information (e.g., by making particular observations, performing particular experiments), or, at the very least, to avoid ignoring pertinent information (e.g., by sticking one's fingers in her ears, or covering one's eyes, etc.). See Booth [2006]. Active testimonial norms might require us to seek information from particular individuals rather than others. In this paper, I investigate only passive testimonial norms. A discussion of active ones is presented in Zollman [2011b]. Second, among passive norms, there might be both norms of *obligation* (e.g. "given evidence E, you should believe p") and norms of *permission* (e.g. "given evidence E, it is permissible to believe p"). See Boghossian [2008]. Arguably, non-reductionists advocate a norm of permission ("you may trust others' testimony in the absence of conflicting information") whereas reductionists advocate norms of obligation ("you must not trust others' testimony in the absence of positive reasons."). In this paper, all testimonial norms are modeled as functions from others' claims to current beliefs; this may obscure the relevant distinction between permissive and obligatory norms, and I would welcome further suggestions concerning how to capture it formally.

[6] In contrast, philosophers have examined the reliability of epistemic norms guiding deduction (e.g., in characterizing valid rules of inference), induction (e.g., in characterizing the conditions under which Bayesian updating, particular belief-revision procedures, etc.

are, to a certain extent, empirical ones, which might require some combination of psychological research (e.g., of common biases and errors of reasoning that might influence speaker reliability), sociological research (e.g., into norms of conversation and truth-telling), and so on. Both projects, however, can also be pursued in the same way that reliability of rules for inductive/statistical inference are evaluated - namely, by modeling. In formal epistemology, statistics, and computer science, researchers develop idealized models of learning from data, and ask, given that data is produced and acquired in a particular way, which rules or procedures reliably lead individuals to develop true beliefs. Further, the conditions under which such rules are reliable can be precisely characterized within a given model, or by modifying the model and characterizing which rules continue to be reliable (i.e., so-called "robustness" testing).

Similarly, one might develop a model of conversational exchanges, and investigate, within such a model, (i) which features influence the reliability of different testimonial norms, and (ii) how reliability changes as those features are modified. This is the aim of this paper.

In Section 1, I develop a formal model of communal learning. The model, I argue, is most appropriate to understanding dissemination of propositional knowledge in *scientific* communities, but I would be happy if the model were applicable elsewhere. In particular, minor modifications to the model might make it appropriate to understanding dissemination of information in everyday conversational exchanges and/or to the transmissions of *behaviors or norms* (in contrast to propositional knowledge).[7]

I then use the model to make precise the concept of a testimonial norm, and to describe six candidate testimonial norms that approximate informal norms such as "believe p if it appears to be the majority opinion", and "believe p if is endorsed by an expert", and "believe p if and only if it appears to be endorsed by a majority of experts." The six norms resemble rules that are endorsed by reductionists and non-reductionists about testimony, and by social epistemologists who advocate "going by the numbers" when deciding whether to accept expert testimony. To be clear, these six norms are extremely simple and naive, and no philosopher, to my knowledge, advocates any of them. However, characterizing the reliability of said norms provides a starting point for characterizing the reliability of the types of norms that are, at least implicitly, under discussion in social epistemology.

lead one to the truth), and inference from perception. Obviously, much of this work has also been carried out by mathematicians, statisticians, computer scientists, and psychologists.

[7] Thanks to Richard Samuels for suggesting the latter interpretation.

In Section 2, I use the model (i) to characterize those features of scientific communities that influence reliability, and (ii) to evaluate the reliability of a wide class of testimonial norms, including the six examples, as relevant features of the community change. I evaluate reliability in three different ways: (1) does employing the testimonial norm eventually lead an individual to develop true beliefs? (2) if so, how *quickly* does the norm lead one to true beliefs?, (3) if error is unavoidable, how often does the norm lead to believing falsehoods? Using these three criteria, I argue that miscommunication and the "social structure" of scientific communities can strongly influence reliability of different testimonial norms. Moreover, changes in norm reliability are often a result of the *interaction* of these two factors.[8] My findings are important because, in traditional, more "individualistic" discussions of the epistemology of testimony, the effects of miscommunication and of social structure are often omitted.

The final section ends with a discussion of limitations of my model and directions for future research.

1 A Model of Communal Scientific Inquiry

In my model, there is some finite set of **questions** that scientists are attempting to answer. Each question has some set of mutually incompatible **answers**, and the scientific community's goal is to find the (unique) correct

[8] Since Kuhn, philosophers of science have argued that the "social structure" of scientific communities can drastically affect the products of scientific research. What exactly "social structure" means has, in many cases, been left intentionally vague. In recent years, philosophers of science have studied several factors that might be called "social structure." For instance, Kitcher [1990, 1995] and Strevens [2003, 2006] study practices for attributing credit for discoveries. Zollman [2010, 2011a] studies the "communicative structure" of scientific communities, i.e., the way in which scientists share and disseminate their findings to one another and through journals. there is also a brief discussion of publications in Weisberg and Muldoon [2009].

In this paper, I will focus on two possible "social structures" of scientific communities, namely whether the community is **interdisciplinary**, in the sense that scientists frequently discuss their research with others outside their field of expertise, or whether it is **insular**, in the sense that scientists rarely discuss their work with individuals outside their field. What I call "insularity" is often called **homophily** among economists and sociologists who employ "network" models like the one I present. Recent work in economics (see Young [2011], and Golub and Jackson [2012]) has investigated the relationship between homophily and speed of learning in such models. Due to constraints on space, I cannot discuss the relationship between my results and those in the existing literature.

answer to each question.[9]

For example, imagine medical researchers are investigating the efficacy of several pills. Each pill is designed to treat a different ailment, and thus, researchers are not interested in *comparing* the effectiveness of the pills. Rather, they only care which pills are salutary and which are harmful. In this example, each question is of the form "Is pill j effective?", and there are two answers: "yes" and "no." Formally, for each pill j, there is some unknown real number e_j that represents the average effectiveness of the pill, where a pill's "effectiveness" is a function of both its side effects and its efficacy in curing the intended ailment. If e_j is positive, then the pill is salutary (on average). Otherwise, the pill is harmful, or at the very least, not beneficial. Moreover, the magnitude of e indicates how harmful or salutary the treatment is. So the formal question is, "Is e_j positive or not?"

I assume that there are discrete stages of time t_1, t_2, and so on, such that, at each stage, each scientist collects **data**. Importantly, such data can be misleading in the short-run. Continuing with the above example, imagine that, on each stage of inquiry, every researcher treats some fixed, finite number of patients with one of the pills and records the results. Those results are the researchers' data. How can such data be misleading? Imagine that the effect of any given pill is probabilistic, and so even if the pill is salutary on average, some patients may react poorly (or not at all) when treated. Thus, when a researcher begins her study, she may observe, as a matter of chance, forty patients who react poorly to the pill, even if it is quite beneficial on average. Similarly for when the pill is harmful.[10]

Here's the complication. I assume that each scientist has a **specialty** (or area of **expertise**), and so each scientist can investigate only one of the questions of interest. In other words, each scientist acquires data that can help her answer only one question. Therefore, she must learn the answers to questions outside her area of expertise by asking others. These modeling assumptions are intended to capture the fact that real researchers' abilities

[9] All bolded terms are defined precisely in Appendix A; they are described only informally in the body of the paper.

[10] In computer simulations described below, I assume there are finitely many pills $1, 2, \ldots, n$. When a patient is treated with pill $i \in \{1, 2, \ldots, n\}$, the scientist observes random effect, which is normally distributed with unknown mean e_i (i.e. the effectiveness of the pill) and unknown variance σ_i^2. The normality assumption is immaterial to all of the results below: similar simulation results are obtained when the agents draw from other types of distributions. Moreover, the theorems below do not depend upon any assumptions concerning the probabilistic process by which data is generated; in particular, sample points are not even assumed to be iid (i.e. they might be correlated, or drawn from different distributions).

are limited due to specialized training, time, and/or financial constraints. Of course, I assume that many scientists might have the same area of expertise. In my running example, imagine each researcher specializes in the study of exactly one pill, and that she treats patients with that pill only. Therefore, she must learn about the efficacy of other pills from the testimony of others.

To model communication, I represent researchers by nodes in a colored, undirected graph like the one pictured below.[11] The colors of the nodes in the graph indicate a researcher's area of expertise, i.e., two researchers share an area of expertise if and only if they are represented by nodes of the same color. For this reason, in my running example, I refer to the various pills as the "red pill", "blue pill" and so on, and the I call a scientist "red" precisely if she studies the red pill.

Edges in the graph represent which scientists communicate with which others. In other words, two scientists can share information if and only if they are connected by an edge. Say two scientists are **neighbors** if they are connected by an edge in the graph; a scientist's **neighborhood**, then, can be defined as the set of all her neighbors.

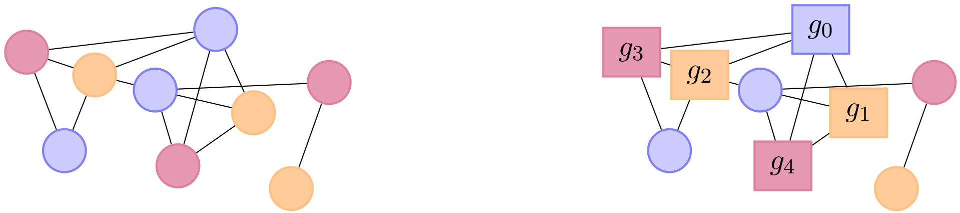


Figure 1: A research network and the neighborhood of g_0 (indicated by squares) in that same network

Not all graphs, however, properly represent scientific communities. Suppose that the graph representing a collection of scientists can be divided into (at least) two sections such that information cannot pass from one section to the other. For instance, see the figure below. In this case, one should not say that the scientists form a single "community," as different parts of the so-called "community" never interact whatsoever. For this reason, I focus exclusively on *connected* networks, which cannot be divided into two

[11] The model here is a member of a large class of models of "network" learning models. See Goyal [2003] for a survey. In this paper, the theorems and summarized simulation results concern undirected graphs, but I have obtained similar simulation results for directed graphs (in which sharing of information may not be symmetric), and it is obvious that all of the theorems hold for directed graphs under minor additional assumptions.

separate parts like below. Formally, define a path to be a sequence of researchers $r_1, r_2, \ldots, r_n$, such that r_1 is r_2's neighbor, r_2 is r_3's neighbor, and so on. A network is said to be **connected** if there is a path between any two researchers.

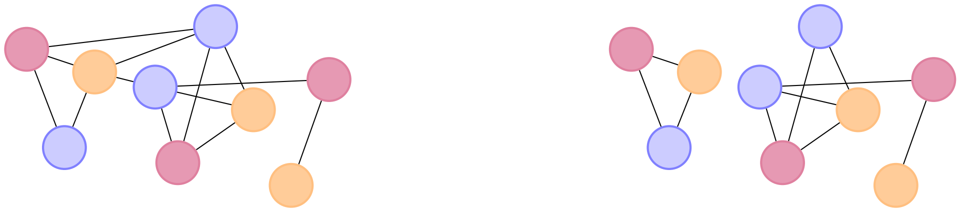


Figure 2: A connected network vs. a disconnected network

Although two neighboring scientists can share information in my model, the type of information they share depends upon their respective areas of expertise. In particular, I assume that two researchers with the *same* specialization can share the *data* they learn, but those with *differing* specialties can only share their beliefs about the *answers* to questions. In the running example, two "red" scientists can communicate how well each of their patients has responded to the red pill. In contrast, a red and a blue scientist can only ask each other "Do you think the red pill is effective?", "Do you think the blue pill is effective?", or even "Do you think the green pill is effective?" and trade answers. That is, scientists with different specializations cannot share their data records, which contain a list of patients and their reactions to the pills, nor can they share their quantitative assessments of *how* effective a pill is.

Why assume that researchers can share information in this limited way? In the real world, scientists must rely on the work and findings of others. However, if scientists could always share and evaluate each other's data, then there would be no such reason to rely on others. The assumption that not all *data* is shared, therefore, is intended to capture the fact that certain "high level" judgments (e.g. is the pill effective?) can be communicated easily even if the data and the methodology for evaluating said data cannot.

But why assume that researchers with the same specialty can share fine-grained information (i.e., data), whereas researchers with different specialties cannot? There are two related reasons. First, it is generally easier for a scientist to understand the findings, methods, etc. of research conducted in her own field than to understand the work of researchers in remote scientific disciplines. For example, theoretical physicists can (sometimes) competently evaluate journal articles in theoretical physics, but can rarely understand more than the abstract and conclusion of a paper in molecular biology. Second, researchers often only read survey or summary articles about work outside their areas of expertise, whereas they often read the journal articles on which summaries are based within their field of research.[12] Thus, even if a researcher could in principle understand work outside her own areas of expertise, she might choose not to do so because of the time-investment it would require to learn more than what is available in survey articles.

Thus far, I have explained two ways in which scientists learn answers to questions in my model, namely, (1) they collect data about a question in their area of expertise, and (2) they learn the answers to questions outside their area of expertise from others. I now explain how my idealized scientists *use* such information to arrive at their beliefs.

Within her area of expertise, a scientist employs a **method** for inferring answers from data. Formerly, a method is just a function from data sequences to answers. I assume that each scientist's method is **convergent**, in the sense that, whatever the truth happens to be, employing the method eventually leads the researcher to the truth with probability one.[13] In other words, given enough data, scientists' methods always output the correct answer to their respective questions.

Three caveats are in order. First, I make no assumptions about how *quickly* such methods find correct answers. For example, suppose that, in my running example, a scientist conjectures that her pill is effective just in case at least half of the patients she has treated have positive outcomes. Now suppose that the pill is effective, but that the first forty treated patients all react poorly to treatment. Such a series of outcomes might be unlikely, but it is possible. Then the scientist's method will lead her astray until she sees a number of positive outcomes, which might take quite a few more observations. In general, I assume that a scientist's method *eventually* will lead her to the true answer to her question, but, there may be no positive number n - no matter how large - such that the scientist is guaranteed to discover the correct answer if her data set contains at least n points.

Second, my imagined researchers do not know when the correct answer has been discovered. There are no "bells and whistles" when the truth is found. In the running example, a scientist may correctly believe that a pill is effective, but she knows that her data might be misleading. Hence, she must entertain the hypothesis that future observations will provide evidence

[12] Thanks to David Danks for suggesting this point.

[13] There are several notions of convergence defined in the Appendix. The theorems in the body of the paper employ the notion of strongly, almost-sure convergence in the Appendix, but similar theorems should hold for the other notions of convergence.

for a different answer than the one she currently endorses.

Third, although I assume researchers eventually discover the truth *in their respective areas of expertise*, no such assumptions are made about finding the truth outside one's specialty. My imagined scientists will need additional rules for learning answers from their colleagues.

In my running example, researchers use statistical methods to evaluate their *quantitative data*. More precisely, a researcher employs a significance test to determine whether her particular pill is effective or not.[14] This part of my model mirrors scientific practice closely, as statistical tests are the trade of most medical researchers and social scientists.

The way in which scientists learn answers to questions *outside their area of expertise* is a little more complex and is explained in greater detail in the next section.

1.1 Testimonial Norms

Recall, from the introduction, that I defined a **testimonial norm** to be a rule for accepting or rejecting the claims of others. In my model, an agent's testimonial norm dictates which answers she believes to questions outside her area of expertise. In this section, I describe six simple testimonial norms. Although the six that I describe are motivated by debates in social epistemology, I should emphasize that I do not think that any philosopher has endorsed norms so simple as the ones below. However, studying these simple norms, I think, can shed light on more complex norms, and many of the results in the next section show that even rather naive-appearing testimonial norms are nonetheless reliable.

Suppose a researcher must decide, on a given stage of inquiry, which answer to believe to a question outside her area of expertise. I call such an agent a **Reidian** if she adopts the opinion of a randomly chosen neighbor in her scientific community.[15] I call her a **majoritarian Reidian** if she adopts the opinion of the majority of her neighbors.

[14] In computer simulations, I assume that treatment effect of pill i is normally distributed with unknown mean e_i (i.e. the effectiveness of the pill) and unknown variance. Researchers employ likelihood ratio tests to determine whether e_i is greater than or equal to zero, or not. In order for said tests to be convergent (i.e. in the limit), the significance of the tests employed is decreased over time at the appropriate rate. See Jeffreys [1998].

[15] I call such an agent a "Reidian" because Thomas Reid is often seen as one of the first philosophers to argue that, in absence of conflicting information, one is justified in believing others' claims. Recall, this is the central thesis of so-called non-reductionists in contemporary epistemology. Of course, the testimonial norm described here is much simpler than what Reid would advocate, as my Reidians randomly chose an agent to trust even when there hear conflicting reports from different neighbors. However, for non-

Notice both types of Reidians ignore their neighbors' areas of expertise. For example, both types of Reidians may trust a red expert concerning questions about the green pill even if they have a green neighbor. In real-world settings, however, individuals often attribute more weight to the opinions of experts, and this reliance on intellectual credentials is often thought to be perfectly reasonable.

To model reliance on experts, I call an agent a **expert truster** (or e-truster, for short) if she adopts the opinion of a randomly chosen expert if one is available, and otherwise, trusts a randomly chosen neighbor. For instance, when a blue e-truster is deciding whether or not the red pill is effective, she asks one of her neighbors who studies the red pill; when she has no red neighbors, then she asks a randomly chosen neighbor. Thus, e-trusters distinguish experts from non-experts. However, when they have expert neighbors, e-trusters trust all such experts equally and distrust all non-experts equally. Call an agent a **majoritarian e-truster** if she adopts the opinion of the majority of her expert neighbors (when she has at least one expert neighbor), and otherwise, adopts the opinion of the majority of all of her neighbors.

Alvin Goldman and others have argued that the norm of e-trusting is unreliable: they claim that agents ought to assess the reliability of both experts and non-experts to determine whom to trust. According to Goldman, an individual can employ various heuristics for evaluating a speaker's reliability. For example, one might attribute greater weight to the opinions of an expert whose opinions are supported by cogent arguments. Unfortunately, not all of the heuristics discussed by Goldman can be accurately represented/captured in my model. Here, I restrict myself to modeling one heuristic for evaluating a neighbor's reliability, which I call **informational proximity**.

The informal concept of informational proximity is best illustrated by examples. I am not a theoretical physicist, nor do I ever communicate with theoretical physicists. However, some of my colleagues, who study philosophy of physics, do in fact speak and collaborate with theoretical physicists. Those colleagues, therefore, are more informationally proximate than I am to current work in theoretical physics. Hence, if a philosophy student is deciding whether to accept my testimony or that of a philosopher of physics when it concerns current work in theoretical physics, she might consider the

reductionists like Reid, recognizing the presence of disagreement is precisely the type of "conflicting information" that can override one's default justification to trust others. I have chosen the name because the testimonial norm here is the most permissive of the six examples I illustrate with respect to whom one is willing to trust.

latter to be more reliable the former because of the informational proximity of philosophers of physics to the facts under investigation.

In my model, the notion of informational proximity can be made precise. Define the **distance** between two researchers to be the shortest path in the undirected graph representing their scientific community. Notice both Reidians and e-trusters ignore informational proximity. For example, suppose an e-truster has two neighbors, neither of which is an expert in the question q. One of the neighbors, however, communicates with a q-expert, whereas the other is three-degrees-removed from the closest q-expert. In such a case, real-world scientists might favor the former's opinion rather than the latter, as the former is closer to the source of reliable information. In contrast, e-trusters in my model ignore informational proximity when they adopt the opinion of a randomly chosen neighbor.

Call a researcher a **proximitist** if, on any given stage of inquiry, she adopts the opinion of the neighbor who is closest to a q-expert when deciding which answer to q to believe; if there are multiple such neighbors, then she chooses one at random. An even more conservative testimonial norm is to poll one's most proximate neighbors; call an agent who follows this norm a **majoritarian proximitist**.

Examples of testimonial norms can be multiplied indefinitely. However, the six norms considered here are important because they differ on several dimensions that have been the focus of debate in social epistemology. By contrasting Reidianism, e-trusting, and proximitism with their majoritarian counterparts, for example, one can investigate the consequences of "going by the numbers" versus those of reliance on one individual. And although no non-reductionist may endorse a testimonial norm so simple as Reidianism, one can investigate the epistemic value of seeking positive reasons to trust a speaker (by employing heuristics like informational proximity) by comparing Reidianism, e-trusting, and proximitism. Perhaps surprisingly, it turns out that Reidians (though not majoritarian Reidians) reliably acquire true beliefs in the absence of miscommunication. More on this later.

Thus far, I have discussed how *individuals* use testimony. However, various testimonial norms interact in interesting ways. For instance, suppose a proximitist and Reidian are neighbors. Further, suppose that the Reidian is the proximitist's neighbor who is uniquely closest to a q-expert. By definition, the the proximist will always trust the Reidian about the correct answer to q. However, the Reidian is willing to trust *all* of her neighbors, including the proximitist. Thus, the two researchers might form temporary "echo chambers", where the proximitist believes some answer a because the Reidian does so, and the Reidian believes a because the proximitist does.

Since testimonial norms might interact in interesting ways when employed in groups, I will consider the reliability of **group testimonial norms** (or GTNs, for short), which assign a testimonial norm to each agent in the network. A GTN is said to be **pure** if every agent is assigned the same norm (e.g., every agent in the network is a proximitist); it is said to be **mixed** otherwise. The next section characterizes the features of scientific communities that affect reliability of GTNs, where reliability is made precise in the four ways discussed in the introduction. I then investigate how reliability varies as those features of scientific communities vary.

2 Reliability

2.1 Convergence

A central goal of scientific inquiry is the discovery of truth. As such, one way to evaluate the performance of GTNs is to investigate, when adopted by a network, whether agents will eventually discover true answers to every question under investigation. Formally, say an GTN is **convergent** if, whatever the truth about the world, when a network adopts said GTN, every researcher will hold only true beliefs given some (potentially very large) finite amount of data. Say a testimonial norm is convergent if any pure GTN consisting of that norm is convergent.

Unfortunately, convergence is insufficient to distinguish among four of the GTNs by the following theorem:[16]

Theorem 1 *In connected research networks, every pure and mixed* GTN *consisting of Reidianism, e-trusting, proximitism, and majoritarian proximitism is convergent. In contrast, there are mixtures of (any subset) of these four testimonial norms with majoritarian Reidianism and/or majoritarian e-trusting that are not convergent.*

Theorem 1 may seem surprising for at least two reasons. First, one may be surprised that Reidianism and e-trusting are convergent norms, as they seem rather naive. Second, one may be surprised that the majoritarian versions of the two norms are not, as one might expect that reliance on the testimony of several individuals is more reliable than reliance on the testimony of one. Understanding (a sketch) of the proof of the theorem can eliminate both surprises. It reveals that the Reidianism, e-trusting, proximitism, and majoritarian proximitism are members of a wide class of convergent norms

[16] Proofs of all theorems are all available in Appendix A.

that satisfy minimal conditions of rationality and descriptive realism. In contrast, majoritarian Reidianism and majoritarian e-trusting are not.

Recall that, by assumptions of my model, all researchers employ convergent methods. That means that, given enough data, every scientist in a community will eventually know the correct answer to the question concerning *her area of expertise.* Thus, to examine which testimonial norms are convergent, it suffices to consider which norms will lead agents to develop true beliefs outside their areas of expertise.

Consider Reidianism first. Outside her area of expertise, a Reidian's beliefs depend entirely upon what her neighbors believed *on the last stage of inquiry.* That is, given the beliefs of her neighbors at time t, one can determine the probability that a Reidian will hold a particular belief at time $t + 1$. Knowing what the Reidian or her neighbors believed before time t is irrelevant. The exact same is true of the remaining five testimonial norms considered here.

For this reason, the six example testimonial norms behave much like **Markov processes**, and in particular, they behave similarly to what are called **absorbing Markov processes**. A common example of a Markov process is the lost tourist. Suppose you are lost in a foreign city and that you cannot tell one street from the next. Because the streets and buildings look similar to you, you may walk in circles without realizing it. So whenever you reach an intersection, you choose a direction at random. In other words, the chances that you will turn left now do not depend upon where you have been, but only upon where you currently are. Now you may choose different directions with different probability (perhaps you really like turning left), and moreover, the chances of choosing any given direction may differ from one intersection to the next. The point is that the chance of turning left at time t depends only upon what intersection you are currently facing. What is the probability that, given an unlimited amount of time, you will eventually find your hotel again?

Under very mild assumptions, it can be shown that the probability is one. For example, it suffices to assume that, at every intersection, there is some non-zero probability that you will choose to move in any given direction. The reason is that there is some path back to your hotel. If there is some non-zero probability of moving in any given direction at every intersection, then there is some probability that you will follow the path back to your hotel exactly. But since you are given infinite time to find your hotel, no matter how small the probability of taking the right path is, you will eventually find it. To see why, consider flipping a coin infinitely many times. Even if the coin is weighted to land heads 99.99% of the time, you should tell expect to observe tails once in ten-thousand flips. An analogous argument holds for the lost tourist finding his hotel.

In my model, Reidians, e-trusters, proximitists, and majoritarian proximitists are much the lost tourist, and the state in which all agents hold true beliefs is much like the tourist's hotel. To see why, consider a network of Reidians. Imagine one of the Reidians - let's call her Jane - is deciding which of her neighbors to trust at time t concerning a question outside her area of expertise. If Jane has an expert neighbor, Jill, then there's some chance that Jane will adopt the Jill's opinion. At time $t+1$, there's some chance that Jane's neighbors will adopt Jane's beliefs. Then at time $t+2$, neighbors of neighbors of Jane may adopt Jane's belief, and so on. In this way, there's some chance that Jill's expert opinion propagates through the entire network (if the network is connected), and this is true at any point in time. So there's some chance that every agent outside of Jill's area of expertise will eventually hold Jill's belief concerning the question of interest.

Since experts eventually hold true beliefs in their area of expertise, this entails that all agents will eventually hold Jill's *true* belief. And once everyone has the same, true belief, the Reidian norm ensures that everyone continues to believe it. Since this is true of every area of expertise, the network must converge. For this reason, the process of belief revisions is said to be **absorbed** by the state in which all agents hold only true beliefs. Similar arguments apply to e-trusters, proximitists, and majoritarian proximitists.

In contrast, majoritarian Reidians and majoritarian e-trusters can behave much like the tourist who always turns left. In other words, there is some chance that majoritarian Reidians and majoritarian e-trusters will never follow the path to true belief. Why? When one employs these two norms, the true opinions of experts can be outweighed by enough contrary opinions of non-experts. Consider a network, like the one below in **Figure 3**, in which there is a community of experts in one field who do not interact with experts in another field. For example, suppose that, at the outset of inquiry, the three neighboring blue experts in **Figure 3** all believe the orange pill to be ineffective. Further, suppose that the three blue experts are all either majoritarian Reidians or majoritarian e-trusters. Because none of these blue experts have a orange expert neighbor, they poll their neighbors to determine the efficacy of the orange pill on each stage of inquiry. Since all three believe the pill to be ineffective initially, they will continue to believe the pill is ineffective forever, regardless of the evidence. So the three scientists will fail to hold true beliefs if the pill were effective, and hence, neither majoritarian Reidianism nor majoritarian e-trusting is convergent.

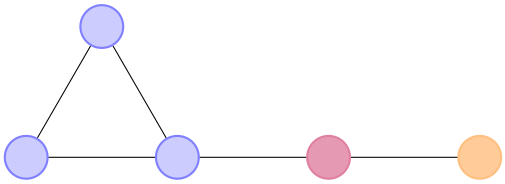


Figure 3: A network in which majoritarian reidians and e-trusters may not converge

The above arguments, however, obviously apply to far more norms than the six defined above. Say an agent's testimonial norm has **finite memory** if there is some (potentially very large) finite number n such that the agent's beliefs depend only upon the last n stages. The six testimonial norms considered here have a memory of length one, but it turns out that increasing an agent's memory has no effect on the above arguments, so long as her memory is finite. Because I am interested in modeling communication among real human beings, it seems natural to consider testimonial norms with finite memory.[17]

So consider an agent employing a testimonial norm with memory of length n. What characteristics of the four convergent norms ensured they eventually found the path to the truth? There were two basic properties at work in the above arguments. Suppose that an agent has been told by all of her neighbors for the last n stages (i.e., for as long as the agent can remember) that φ is the correct answer to a question outside her area of expertise. Intuitively, the agent should believe φ with probability one, as she has no evidence to the contrary. Call a testimonial norm with this property **stable**. Stability ensures that consensus perpetuates itself, and in particular, it ensures that if all agents hold true beliefs, then the network has been "absorbed" into the state of all true belief.

The second property ensures that experts' true opinions can propagate through the network with some positive probability. Fix an agent A and a

[17] Two notes are in order. First, one might question whether having finite memory is best modeled by assuming the agent can remember the *most recent* n stages, rather than remembering n stages in *total* (not necessarily consecutive, or most recent). I believe that all of the theorems discussed here continue to hold under this alternative assumption, but more work is necessary. Thanks to Patricia Rich for suggesting this alternative assumption. Second, although I assume testimonial norms have finite memory, I do not assume anything similar about agents' methods for answering questions in their own areas of expertise. In fact, for many of the paradigmatic questions that I use as examples (in the Appendix), agents' methods must have unbounded memory in order to be convergent. This is an unfortunate combination of assumptions, and further work is necessary to model descriptively feasible methods.

question q outside the A's area of expertise. Say A's testimonial norm is **q-sensitive** if there is some (potentially very small) non-zero probability ϵ and some set N of A's neighbors such that (i) N is a subset of A's neighbors who are most proximate to a q-expert, and (ii) if every agent in N believes that φ is the correct answer to q, then A believes φ with probability at least ϵ. Say A's testimonial norm is sensitive if it is q-sensitive for all questions outside her area of expertise. Then we obtain:

Theorem 2 *Any mixed or pure* GTN *is convergent if it consists of finite-memory norms that are sensitive and stable.*

So there is nothing special about the four convergent norms. They are members of an extremely wide class of convergent testimonial norms satisfying basic requrirements of rationality (i.e., stability and sensitivity) and realism (i.e., finite memory). In contrast, majoritarian Reidianism and majoritarian e-trusting are not sensitive norms. This does not entail that they are not convergent (as the converse to the above theorem is false), but it does explain in what ways these two norms differ from the remaining ones.

The philosophical upshot of the above theorems is that, if reliability is understood solely in terms of the eventual acquisition of true beliefs, then there is no epistemic difference among a wide class of testimonial norms.[18] In particular, the decision to adopt a "reductionist" norm, which might require one to find positive reasons to trust a speaker, versus a "non-reductionist" norm, which might permit individuals to trust others in the absence of defeating conditions, is unimportant as long as the norms in question are sensitive and stable.

The above two theorems, however, seem to depend crucially upon the assumption that agents have infinitely long to acquire true beliefs. In particular, I have neglected the *speed* with which agents learn. One might wonder, "if reliability is understood in terms of *quick* convergence to the truth, then are there any differences among the four convergent testimonial norms?" Surprisingly, the answer is "no," and this is the topic of the next section.

[18] A similar conclusion holds for methods for inferring theories or hypotheses from data. Reichenbach justified his "straight-rule" by arguing that, in the long run, the straight rule outputs increasingly accurate answers over time. Carnap [1996] criticized Reichenbach because, in many paradigmatic applications of the straight rule, there are infinitely many methods for inferring hypotheses from data that also exhibit long-run reliability.

2.2 Speed of Convergence

Define the **convergence time** of a network to be the number of stages elapsed before every agent in the network holds true beliefs and will continue to hold such true beliefs indefinitely. Thus, the first way of evaluating various GTNs is to consider the question, "which GTNs minimize average convergence time?"

To evaluate the effect of GTNs on convergence time, I simulated the running example of the model described in the previous section.[19] First, I randomly generated approximately 4500 undirected graphs consisting of between 50 and 100 agents. If a graph was disconnected, it was removed from the data set because none of the GTNs discussed above is convergent in disconnected networks.[20] To model the fact that communication is limited, I generated only graphs in which researchers could communicate with at most 10% of the other agents.

Other than ensuring the network was connected and that the number of edges was limited, I made no attempt to generate "realistic" networks that resemble actual scientific communities; any network satisfying these two constraints was possible, including ones in which researchers form a line **Figure 4** and in which an expert in question q is more informationally proximate to an expert in $q' \neq q$ than she is to any expert in her own field. Below, I describe the results of the simulations when more realistic networks were considered.

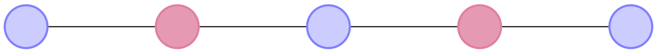


Figure 4: Unrealistic Networks: A Line Network

Equal numbers of agents were assigned one of five areas of expertise (i.e. the study of one of five pills). The network was then assigned either one of the four pure GTN or one of 1000 mixed GTNs. At each stage of the simulation, the researchers acquired data and updated their beliefs as described in the previous section. A simulation was stopped when all agents' beliefs were true for ten consecutive stages of inquiry, and the tenth to final stage was assumed to be the convergence time of the network.

[19] I employed the NetLogo programming language for all computer simulations. For information concerning NetLogo, see Wilensky [1999]. The code for the simulations can be found online at `http://www.andrew.cmu.edu/user/conormw/Papers.htm`.

[20] Connectedness is sufficient but not necessary for consistency. Call a network **expert connected** if every connected component contains at least one expert of each type. Note that every connected network with at least one expert of each type is expert connected. Then every pure or mixed GTN consisting of the four typical norms is consistent if and only if the network is expert connected. See Appendix A.

What effect(s) do GTNs have on convergence time?[21] The answer: essentially none. Except in the "easiest" problems (where "ease" will be defined shortly), there is no statistically significant effect of choice of pure GTN on convergence time. In other words, populations of exclusively Reidians, populations of exclusively proximitists, and so on, all converge at the same rate on average. In "easy" problems, Reidians converged at a rate slower than the remaining pure GTNs on average, but there is no significant difference among the remaining pure GTNs.

The results for mixed GTNs are similar. When the problem is sufficiently hard, there is no statistically significant difference among the 1000 mixed GTNs, and when the problem is easy, mixtures containing more Reidians converge at slower rates.

On first glance, these results may seem surprising and might even seem like evidence that something is very wrong with my model. In order for Reidians to converge, a very specific sequence of events need to occur: when deciding which answer to believe for question q, agents must first trust their q-expert neighbors, and then neighbors of said agents need to trust those who received information from q-experts, and so on. Although I have argued, in the previous section, that this improbable sequence of events will transpire at some point, it is not guaranteed to happen quickly. In contrast, in a network of proximitists, true belief propagates from experts to non-experts as soon as it is discovered. So how can Reidians and proximitists converge, on average, at roughly the same rate in difficult problems?

I claim these findings are intuitive and reveal a robust pattern in the history of science when the notion of difficulty is properly understood. To understand the notion of difficulty, imagine that some scientific community is faced with two different questions. Suppose there are two possible answers to each of the two questions. For simplicity, call the answers to first question T_1 and T_2, and call the answers to the second T_3 and T_4. Suppose that T_1 and T_2 are mutually contradictory (i.e. if T_1 is true, then T_2 if false, and vice versa), so one cannot reasonably endorse both. Similarly for T_3 and T_4.

With respect to question one, suppose that scientists have designed an experiment that will, with high probability, rule out either T_1 or T_2, depending on which was false. In contrast, there is no "crucial" experiment to answer the second question: theories T_3 and T_4 entail that similar observations would be made in almost all experimental settings. Hence, without

[21] The effects of pure GTNs on convergence time were compared using a one-way random effects analysis of variance. Sample statistics and an explanation of the statistical methods employed to analyze simulation data is available in Appendix B.

sufficiently large samples, precise measuring instruments, and/or ingenious experimental design, researchers will be unable to determine which of the latter two theories is true. Intuitively, it seems that the second question is more difficult to answer than the first. In general, we can (roughly) define the **difficulty** of a question to be the degree of precision that is required to distinguish among all possible answers, where precision may be gained by acquiring more data, employing better measurement tools, and so on.

In the running example of my model, this notion of difficulty can be made precise. If the pills in question are very effective or very harmful, then researchers need observe only a few patients to learn this fact. However, if the pills are only slightly salutary or harmful, then much larger samples are necessary. Therefore, in simulations, I used the absolute value of the effectiveness e_j of pill j as a measure of difficulty.

Clearly, when agents are faced with more difficult questions, they learn more slowly *within their area of expertise.* However, the time they take to communicate their findings to those outside their area of expertise is the same as if the question had been easy. It should now be clear why testimonial norms have no significant effect on convergence time when the questions of interest are difficult. Testimonial norms affect only the *dissemination* of information to non-experts, whereas methods are responsible for the time it takes experts to *discover* the true answers. As the questions under investigation become more difficult, the time require to disseminate information is dwarfed by discovery time. Hence, testimonial norms have only a negligible effect on convergence time when questions are difficult.

This is not an artifact of my model, but is also readily observable phenomenon in the history of science. Consider any difficult scientific undertaking - for example, understanding the principles of flight. Before the Wright brothers, human beings had attempted to engineer airplanes for millennia. So the discovery of principles of flight took at least a few thousand years. In contrast, once the first airplanes had been constructed, the engineering knowledge spread around the globe in a matter of a few years. The time to disseminate such knowledge, therefore, was minuscule in comparison to the time it took to gain it in the first place.

So we have seen that infinitely many GTNs are convergent, and when the questions under investigation are difficult, there is no significant difference among GTNs in terms of *quickness* of convergence to truth. Since science is a difficult enterprise, one might conclude that choice of testimonial norm is epistemically irrelevant for scientists. This conclusion is hasty. Although there are several idealizations in my model, there are three that have been neglected thus far that deserve close scrutiny.

First, I have assumed that that researchers can adopt the *beliefs* of their neighbors. But that is equivalent to assuming that agents never misspeak or misinterpret others' claims. Do the performances of the various GTNs change when miscommunication is possible? Second, I have evaluated the performance of GTNs over *all* possible networks. However, many networks do not represent the types of communities formed by real world scientists. Do the relative performances of GTNs change in more realistic networks? Finally, because e-trusting and proximitism require a scientist to identify which of her neighbors are experts (and even those most proximate to experts!), I have assumed that scientists can evaluate the knowledge and ability of researchers outside their specialization. However, identifying experts is often very difficult. Do the above results change when agents misidentify experts? These three questions are the subject of the next three sections.

2.3 Miscommunication

Miscommunication is an unavoidable feature of human conversation. Speakers make (grammatical, vocabulary-related, etc.) errors that result in ambiguity and/or unintended meanings, and listeners may misinterpret or misunderstand what speakers say. Anecdotally, misunderstandings seem fairly common in academic communities when researchers in one field try to share their findings with non-experts. Does such miscommunication affect the reliability of various testimonial norms?

The answer is "yes", and it is helpful to consider the relationship between honest miscommunication and dishonesty to see why.[22] Intuitively, in communities in which lying is widespread, relying on the testimony of others will be less reliable than in communities in which individuals are honest. The reason seems to be fairly obvious: when a lie is successfully told, the speaker believes some proposition φ, and the listener is led to believe a proposition ψ that is incompatible with φ. If the speaker has a true belief, then the listener will have a false one. Notice the exact same story is true in many circumstances of miscommunication. When miscommunication occurs, the speaker has one belief, and the listener another. Often times, the former is true and the latter is false. So if dishonesty can undermine the reliability of testimonial norms, then one should expect miscommunication to do the same.[23]

[22] Thanks to John Greco for this suggestion.

[23] Of course, the degree to which dishonesty and miscommunication affect reliability might be different. Here are two candidate differences between dishonesty and miscommunication. First, if a speaker lies and she is asked to explain or clarify what she has

This may seem obvious, but much of the literature in the epistemology of testimony has neglected miscommunication and focused almost exclusively on dishonesty. Since both might affect the reliability of testimonial norms, for the remainder of the paper, I will say that **miscommunication** has occurred when a speaker believes φ, and as a result of the speaker's testimony, an agent comes to believe some proposition ψ that is incompatible with φ. In the general model I have presented, this broad notion of miscommunication might be made precise in any number of ways. Hence, for the remainder of the paper, I will investigate the effect of miscommunication only within the running example.

In the running example, I assume there is fixed probability $\epsilon < \frac{1}{2}$ with which, on any given stage of inquiry, agents lie to one another, misunderstand one another, misspeak, etc. That is, if g believes the red pill is effective and a neighbor n asks g her opinion, then n will believe that g reported the red pill to be *in*effective with probability ϵ. I assume ϵ is less than a half because I am uncertain whether assertions can have a fixed meaning if miscommunication is more probable than the flip of a fair coin. None of the results below rely on the fact that ϵ is the same for all agents; nor do they depend upon the fact that misspeaking, misinterpretation, and/or dishonest communication are all lumped into one general notion of miscommunication. These assumptions are made for simplicity of calculations and proofs only.

As in the previous section, one can ask, "which of the pure and mixed GTNs discussed are convergent when miscommunication is present?" The answer: none.

Theorem 3 *Suppose there is some fixed, non-zero probability of miscommunication. Then no pure or mixed* GTN *consisting of the above six testimonial norms considered above is convergent.*

The justification for the above theorem is simple. Because there is some fixed probability (however small) that agents will misspeak or misinterpret others, there is always some finite probability that, even when all of one's neighbors hold true beliefs, one will accidentally misinterpret their claims.[24]

said, then it seems likely that she will reiterate the same false statement. When miscommunication occurs, clarification might eliminate the transmission of false beliefs. So, in *repeated* conversational exchanges, errors due to miscommunication might be mitigated more easily than errors due to dishonesty. Second, when a speaker successfully lies to a listener, the latter will come to believe something the former regards as false. If the former had a true belief, then the latter will have a false one. When miscommunication occurs, listeners may arrive at true beliefs accidentally.

[24] The assumption that the probability is fixed is sufficient, but not necessary for the

If the possibility of miscommunication is high, researchers' beliefs may fluctuate chaotically. It is possible for every agent to hold only true beliefs on one stage, and only false beliefs on the next.

One might worry that the failure of the above norms to converge is again an artifact of their naiveté. For instance, the six norms considered here all require an agent to change her beliefs in light of what she learns from he neighbors on the *most recent stage* of inquiry. One might worry that this tantamount to assuming that, if a professor delivers the same lecture one thousand times to the same class, then students may change their beliefs if she misspeaks on the thousandth iteration. A more realistic assumption is to alter the above testimonial norms so that agents *remember* past. For instance, one might try to model agents who only alter their beliefs only when the underlying "signal" from a neighbor has changed. That is, agents change their beliefs only when they cannot attribute the differences between past and current utterances of their neighbors to miscommunication. However, if agents can remember only *finitely* many stages of inquiry into the past, then even the most ingenious testimonial norms are subject to the same argument as above: no GTN converges.

Theorem 4 *Suppose there is some fixed, non-zero probability of miscommunication. Then no pure or mixed* GTN *consisting of norms with finite memory is convergent. In particular, none of six testimonial norms considered above is convergent.*

The above argument only shows that agents might *occasionally* believe false statements. Perhaps it is wrong, then, to demand that testimonial norms converge. Rather, one should be interested in testimonial norms that *minimize error*. When convergence is possible, minimizing error in the long-run should require one to employ convergent norms, and when convergence is not possible, the criterion of error minimization can help to distinguish among rival testimonial norms.[25]

How should we calculate error? Imagine taking a snapshot of all agents' beliefs on a given stage of inquiry. Given the snapshot, one can calculate

result. As long as the frequency of miscommunication does not decrease too quickly, then convergence is prevented (by the Borel Cantelli Lemma).

[25] Of course, it has been long recognized that seeking truth and avoiding error are two different epistemic goals that may pull one in different directions. See James [1896]. Because agents in my model cannot abstain from belief, and because answers to questions are incompatible with one another, avoiding error in my model requires having true beliefs. Future work ought to relax these assumptions and consider testimonial norms that allow agents to abstain from belief and to believe disjunctions of answers.

the proportion f_n of all agents' beliefs that are false. Given a particular set of questions and a model of how data is generated, one can (in theory) calculate the *expected* number of false beliefs in a network $e_n = E[f_n]$. Call e_n the **error rate** of the network **on stage** n. For many testimonial norms, the error rate e_n fluctuates greatly from one stage to the next. Luckily, the six testimonial norms that we have considered are not of this sort:

Theorem 5 *Suppose there is some fixed probability of miscommunication. Then every pure and mixed* GTN *of the six testimonial norms above converges to some fixed error rate. That is,* e_n *approaches some fixed* e *value as* n *approaches infinity.*

Call the fixed value e the **error rate** of the network (simpliciter). Hence, one can compare the performance of various GTNs by comparing their respective error rates. Surprisingly, when miscommunication is possible, the link between testimonial norm and truth becomes clearer. Simulation results show that, for all problem difficulties and all networks, the error rate of populations of Reidians is on average greater than that of e-trusters. Populations of e-trusters err more often than do proximitists, who in turn, err more than majoritarian proximitists. The statistical tests supporting these claims are summarized in the Appendix.

Although Reidians err more often than do e-trusters and proximitists, it turns out that the error rates of various testimonial norms differ widely across different "network structures." This is the subject of the next section.

2.4 Network Structure

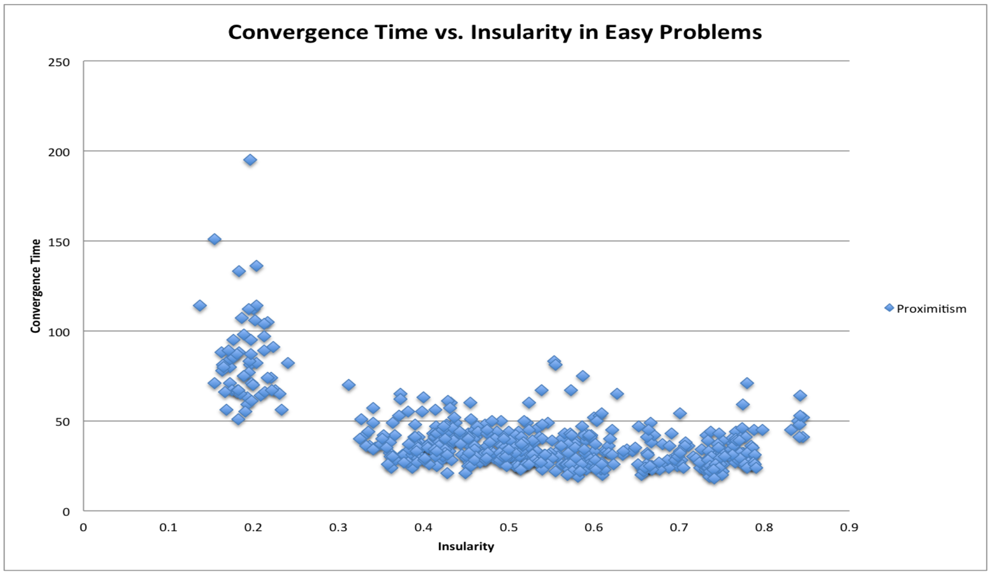
Thus far, I have argued that, if communication is perfect and one's goal is the quick acquisition of true beliefs, then choice of testimonial norm is irrelevant. In contrast, if miscommunication is present, then different testimonial norms have differing rates of error. In analyzing the simulation results, however, I have made no attempt to distinguish realistic network structures from mathematically possible, but highly unrealistic, ones. For instance, in my simulations, I could have (but did not) randomly generated a "line" network like in **Figure 4**, which obviously does not resemble any real-world scientific community. Are there any features of real scientific communities that might affect the performance of GTNs?

One way in which academic communities are unique is that they are often divided into *research units*. Roughly, a research unit is a collection of individuals who (i) have similar research programs and (ii) communicate with one another frequently. In some cases, a research unit may be a particular academic department at a particular university. In other circumstances, research units are comprised of academics who live in different parts of the world, but still collaborate on papers, read each others' independent work, and so on.

Because the graphs in my model represent scientists' expertises and how they share their findings, they are perfectly suited to model research units. In my model, I represent research units by collections of agents who (i) share an area of expertise and (ii) are highly connected by a collection of edges. Further, the *degree* to which scientists form a research unit can also be made precise. Given an agent g, define the **insularity of g's communication** of g to be proportion of agents in g's neighborhood who share g's expertise. Define the **insularity** of a network to be the average insularity of all agents.

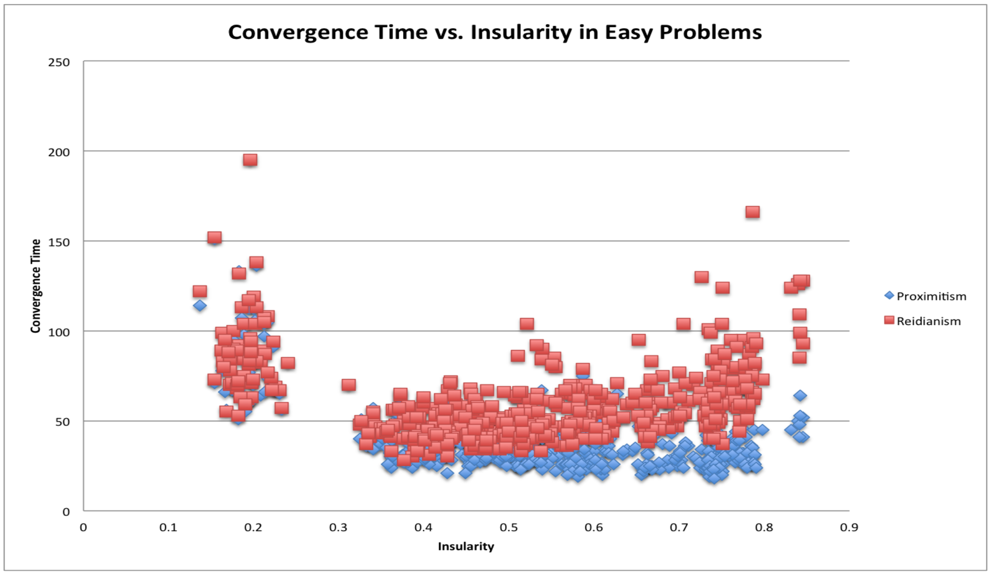
Intuitively, insularity so-defined seems to be both desirable and dangerous. On one hand, researchers with similar expertise ought to communicate and collaborate as frequently as possible; so insularity is desirable. On the other hand, when academic communities become too insular, there is a chance that one research unit completely isolates itself, thereby failing to share its own findings or draw upon the work of others. So too much insularity is harmful.

Again, these intuitions are captured by my model. For the moment, let us ignore miscommunication once again. Below is a graph illustrating the relationship between the convergence times of several networks of proximitists and the insularity of those networks. The graph shows that proximitists converge more quickly (on average) in more insular networks. However, recall that, in my simulations, I discarded all data generated by disconnected networks, as such networks may not converge to true belief at all. Together, the graph and the omitted data capture the above intuitions about insularity precisely: if quick convergence is the goal of science, then increasing the insularity of a network seems to be desirable until it fractures the network into smaller, isolated (i.e. disconnected) research communities.



The reason that insularity decreases convergence time is fairly simple. Recall that scientists with the same expertise can share data in my model. More data allows the researchers to more quickly ascertain the true answer in their own areas of expertise, and therefore, decreases convergence time.

The above graph depicts a fact about networks of proximitists only. Similar graphs are obtained for the other three convergent (pure) GTNs, *with one important exception.* Recall that, when the underlying problem is easy, populations of Reidians converge less quickly than do communities employing the other three pure GTNs. It turns out that Reidians do not *always* underperform e-trusters, proximisits, and majoritarian proximitists: they only do so when the underlying network is insular. Below is a graph illustrating this fact. At first, increasing the insularity of a network increases the speed at which Reidians converge, until a critical point is reached at which their speed slows significantly.

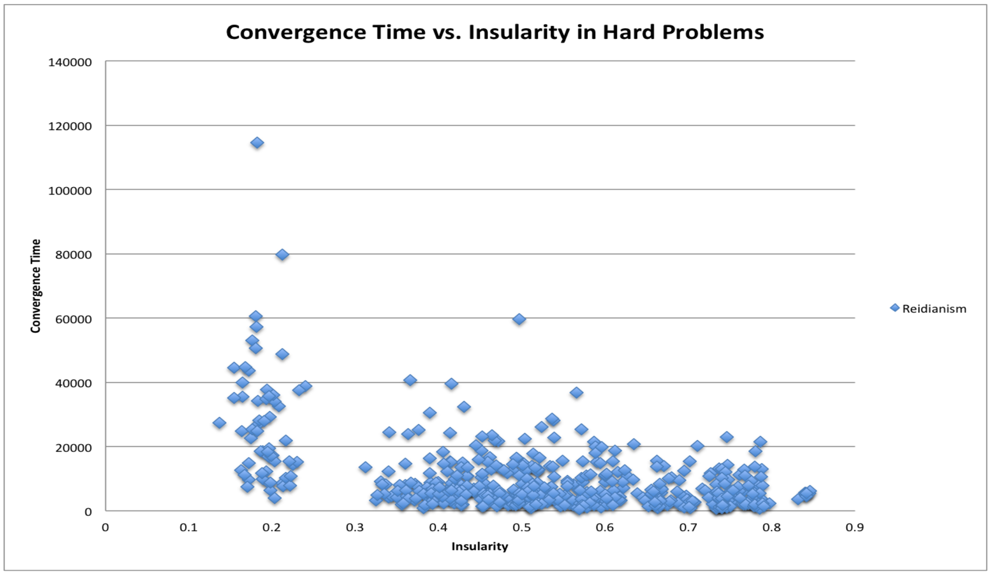


To explain the relationship in the above graph, it is helpful to recall that convergence to true belief is a two-stage process consisting of *discovery* and *dissemination*. In the discovery stage, scientists acquire data that allows them to ascertain the truth in their respective areas of expertise. In the dissemination stage, the true views of experts propagate through the network due to the GTNs. Recall that, when convergence is viewed in this way, GTNs play a significant role only in the dissemination stage.

For populations of e-trusters, proximists, and majoritarian proximitists, the views of the expert, research units propagate quickly through the entire community, even when the network is insular. So the dissemination stage of convergence is always quick. In contrast, for Reidians, dissemination takes more time on average when the network is insular, as scientists with different areas of expertise become more and more isolated from one another. Thus, all other things being equal, one should expect Reidians to converge more slowly precisely when the dissemination stage constitutes a significant chunk of the convergence process.

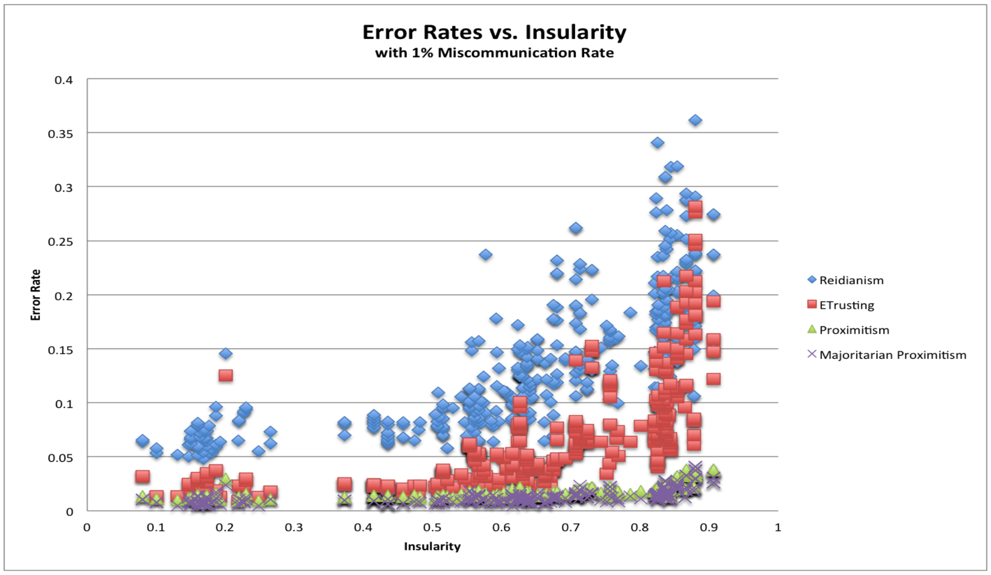
Now, when the problem is easy, it is solved quickly by experts. As a result, the discovery and dissemination stages are approximately equal in length for populations of e-trusters, proximists, and majoritarian proximists. In contrast, for Reidians, the second stage takes more time in insular networks, resulting in a statistically significant increase in average convergence time.

In contrast, when the problem is difficult, the discovery stage is considerably longer, as research units take greater time to find the truth. So the length of the discovery stage dwarfs that of the dissemination state, regardless of which GTN is employed. Recall, this is why there is no statistically significant difference among the convergence times of the GTNs when the underlying learning problem is difficult. Moreover, it also explains the relationship illustrated in the graph below, namely, that when the problem is difficult, more insular networks of Reidians converge more quickly on average. Why? Since increasing insularity quickens the discovery stage, and since the discovery stage is significantly longer when the question is difficult, the gains in discovery speed outweigh the losses in dissemination speed.



When miscommunication is present, a similar result holds when one considers the time it takes a network *to converge to its error rate*. Because agents employ convergent methods in my model, the error rate of a network is determined entirely by the number of false beliefs agents hold with respect to questions outside their area of expertise. During the discovery stage, the frequency of false beliefs in a network will typically be higher than the (asymptotic) error rate, as by definition, during the discovery stage agents may have false beliefs in their own area of expertise. It follows that shortening the discovery stage shortens the time until a network converges to its error rate. Since insular networks have shorter discovery stages, they will also typically converge to their error rates more quickly.

What is the relationship between the *magnitude* of error rates and network structure? One might expect that insular networks also have lower error rates. In fact, the opposite is the case. Below is the graph that shows that, for each of the four pure, convergent GTNs, more insular networks typically have higher error rates. However, the rate at which error rates increase as a function of insularity differs among the four GTNs. In the presence of miscommunication, both radical and e-trusters quickly become unreliable as insularity increases, wheras both proximists and majoritarian proximitists have slower growing error rates.



Hence, the possibility of miscommunication raises a dilemma. On the one hand, more insular networks have higher error rates. On the other hand, such networks also converge more quickly to said error rates. So there is a direct trade-off between speed of acquisition of (mostly) true belief, and the number of true beliefs (on average) in research communities.[26] How such a trade-off ought to be handled, I think, will depend largely on the questions of interest and the degree to which getting information quickly is valued.

The reason that insularity increases the error rates of all the testimonial norms under consideration is fairly easy to explain. Essentially, error

[26] [Zollman, 2011a] finds a very similar trade-off between speed and reliability in a different model of scientific inquiry. This is evidence that the phenomenon (i.e., the trade-off) is robust under varying modeling assumptions.

rates are a consequence of a “telephone-game effect” that arises when miscommunication is present. When an agent learns a fact first-hand from an expert, there is only a small chance of miscommunication. When an agent learns a fact second-hand, the chance that miscommunication has occurred is higher: not only is there a chance of miscommunication between an agent and her informant, but also there is a chance of miscommunication between the informant and the expert from which she learned the fact. So as the length of the informational path between an agent and expert is increased, so is the chance that miscommunication has occurred somewhere along the way. When a network is insular, informational paths between two experts in different fields are generally longer, and hence, error rates are typically higher. This suggests that to minimize error rate and ensure high speeds of convergence, ideal network structures ought to balance insularity and average path-length between agents; so-called “small worlds” networks often have this property precisely.

One might worry that the simulations and analytic results above paint Reidianism in poor light only because I have assumed that e-trusters and proximitists can do the impossible - namely, they reliably identify experts in subject matters about which they know nothing. How do e-trusters and proximitists perform when they make mistakes concerning who is an expert? This is the subject of the next section.

3 Identifying Experts

Thus far, I have argued for three theses. First, in the absence of miscommunication, a large class of testimonial norms are convergent. Second, in difficult scientific enterprises, many testimonial norms converge at the same speed. Thus, third, only when miscommunication is present can one distinguish between the value of differing testimonial norms, and it is the interaction of miscommunication and social structure that produces such differences in reliability. However, one might question whether the above results are robust, as I have assumed that scientists can infallibly recognize experts outside their own area of expertise. This is a highly dubious assumption, as identifying experts may require precisely the knowledge that expert testimony is intended to supply. The purpose of this section is to show that, in fact all, three theses are robust even when agents are unreliable in identifying experts.

With respect to the first thesis, one might ask, “in the absence of miscommunication, is there any guarantee that e-trusting and proximitism are convergent if it is possible to misidentify experts?" To answer this question, suppose that each scientist can correctly identify experts within her own field, but that, in different subject matters, she is fallible with respect to expert identification. Specifically, assume that on the n^{th} stage of inquiry, every scientist correctly identifies which of her neighbors are experts (or are most proximate to an expert) with some probability $p_n < 1$.

Notice, the probability p_n might be very low: agents may only correctly recognize experts one in a million times. Moreover, agents may become more or less reliable in identifying experts as time passes. Nonetheless, under a fairly weak assumption, the first thesis remains true even when experts may be misidentified.[27]

Theorem 6 *Suppose that the infinite sum $\sum_{n \in \mathbb{N}} p_n^k$ is infinite for all natural numbers k. Then any mixed* GTN *consisting of Reidianism, e-trusting, proximitism, and majoritarian proximitism is convergent. In contrast, for some series of probabilities p_n satisfying the above conditions, there are mixtures containing majoritarian Reidianism and majoritarian proximitism that are not convergent.*

How strong is the assumption that $\sum_{n \in \mathbb{N}} p_n^k$ is infinite for all natural numbers k? Not very. For example, the assumption is satisfied when p_n is any constant, no matter how small. So if there is any lower bound on agents' chances of misidentifying experts (say one in a billion), then a wide variety of testimonial norms are convergent. Perhaps more surprisingly, the assumption is met even when p_n is equal $\frac{1}{\log n}$, which represents the situation in which agents' ability to identify experts (or those most proximate to experts) *approaches zero* as inquiry progresses. In other words, so long as expert misidentification does not increase quickly, many GTNs will converge nonetheless.

What about the second thesis? In the presence of expert misidentification (but no miscommunication), do different GTNs converge at differing rates? Again, the answer is "not in difficult problems", and the same argument as before is still applicable. Recall that convergence is a two-stage process consisting of discovery and dissemination. In the discovery stage, scientists identify the correct answers to the questions in their respective areas of expertise. In the dissemination stage, those answers propagate through the community.

[27] One might ask whether it is crucial to assume (i) that p_n is the same for all agents in a network and (ii) that scientists' abilities to recognize experts does not depend upon the difficulty of the subject matter. It is not: the assumptions are made for convenience only.

Now recall that I have assumed that scientists can correctly identify experts within their own area of expertise. This entails that expert misidentification affects only the speed of dissemination. Because dissemination has a negligible effect on convergence speed in difficult inquiries, it follows that testimonial norms have an insignificant effect on convergence speed, even when experts may be misidentified. Of course, in easy problems, expert misidentification slows the speed with which e-trusting, proximitism, and majoritarian proximitism converge.

What happens when expert misidentification and miscommunication are both present? Under one additional assumption, it turns out that one can calculate error rates of GTNs as before:

Theorem 7 *Suppose that there is some finite chance of miscommunication and that the probability of expert misidentification is constant (i.e., there is some finite q such that $p_n = q$ for all stages of inquiry n). Then any mixture of the above six testimonial policies approaches a fixed error rate.*

Therefore, if one assumes that the probability of expert misidentification is constant, then GTNs can be ranked by comparing error rates, just as I did before.

4 Conclusions and Future Research

Recall, my goal is this paper was twofold: (i) to identify those contextual features that influence the reliability of testimonial norms, and (ii) to assess the reliability of various testimonial norms as those contextual features vary. I have argued that miscommunication and the degree of insularity of scientific communities are two such contextual features. As either feature is increased, so is the error rate attributable to testimonial norms. However, more insular networks typically converge to (mostly) true belief more quickly, and hence there is trade-off between speed of learning and error. It is an open question whether there are any testimonial norms for which error rates *decrease* as the network becomes more insular.

The paper leaves open a number of questions and areas for future research. One obvious project is to investigate the reliability of more realistic testimonial norms. The six norms that I have studied are simplified versions of ones that have been suggested by epistemologists, but clearly, the types of testimonial norms are employed by real human beings are far more complex and nuanced. This paper has provided the concepts and tools for studying such complex norms, but the work remains to be done.

There are also a number of idealizations and/or limitations of my model that deserve further investigation. I will mention three. First, in my model, network structure is *static*, but clearly real scientific communities change: new research collaborations are formed and old ones break apart. A more realistic model would allow the underlying network structure to change. Such dynamic scientific communities raise a number of new questions. How should the underlying graphical structure representing scientific communities evolve to mirror the real-world dynamics of scientific communities? Is there a way to extend the concepts of "insularity" and "informational path length" to dynamically evolving networks?

Second, in my model, areas of expertise are disjoint and unrelated: one agent's findings are useless to researchers with different expertises. In real scientific communities, a mathematician's findings can help physicists solve problems; physicists' models can be applied to economic phenomena; economists' techniques can be applied in biology, and so on. A more realistic model, therefore, would be capable of representing the complex collaborative relationships among different academic disciplines.

Finally, in my model, agents exchange answers to questions without providing reasons for their opinions. A more realistic model of communication might represent the exchange of *arguments* as well as answers. How one should define and model testimonial norms when agents exchange reasons remains an open question, but it is crucial to investigating the reliability of the more realistic, complex norms that humans actually use.

References

Jonathan E. Adler. Testimony, trust, knowing. *The Journal of Philosophy*, 91(5):264—275, 1994.

Paul A. Boghossian. Epistemic rules. *The Journal of philosophy*, 105(9): 472—500, 2008.

Anthony R. Booth. Can there be epistemic reasons for action? *Grazer Philosophische Studien*, 73(1):133—144, 2006.

Tyler Burge. Content preservation. *The Philosophical Review*, 102(4):457—488, 1993.

Rudolf Carnap. The aim of inductive logic. *Logic, probability, and epistemology: the power of semantics*, 3:259, 1996.

C.A.J. Coady. Testimony and observation. *American Philosophical Quarterly*, 10(2):149—155, 1973.

R. Durrett. *Probability: Theory and Examples.* Cambridge University Press, 2010.

Richard Foley. Universal intellectual trust. *Episteme: A Journal of Social Epistemology*, 2(1):5 — 11, 2005.

Elizabeth M. Fricker. Against gullibility. *Synthese Library*, pages 125—125, 1994.

Elizabeth M. Fricker and D.E. Cooper. The epistemology of testimony. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 61:57—106, 1987.

Alvin I. Goldman. Experts: Which ones should you trust? *Philosophy and Phenomenological Research*, 63(1):85—110, 2001.

Benjamin Golub and Matthew O. Jackson. How homophily affects the speed of learning and Best-Response dynamics. *Forthcoming in Annals of Economics and Statistics*, 2012.

Sanjeev Goyal. *Learning in Networks: a Survey.* University of Essex, 2003.

Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability.* American Mathematical Society, 1997.

William James. The will to believe. *The New World*, 5:327—347, 1896.

Harold Jeffreys. *Theory of Probability.* Oxford University Press, USA, 1998.

Philip Kitcher. The division of cognitive labor. *The Journal of Philosophy*, 87(1):5—22, 1990.

Philip Kitcher. *The Advancement of Science: Science Without Legend, Objectivity Without Illusions.* Oxford University Press, USA, 1995.

Jennifer Lackey. Testimonial knowledge and transmission. *The Philosophical Quarterly*, 49(197):471—490, 1999.

Jennifer Lackey. Testimony: Acquiring knowledge from others. *Social Epistemology: Essential Readings, ed. Alvin I. Goldman & Dennis Whitcomb. Oxford: Oxford University Press*, pages 314—337, 2011.

Keith Lehrer and Carl Wagner. *Rational Consensus in Science and Society: A Philosophical and Mathematical study*, volume 24. D. Reidel, 1981.

John L. Pollock. Epistemic norms. *Synthese*, 71(1):61—95, 1987.

Peter Spirtes, Clark N. Glymour, and Richard Scheines. *Causation, prediction, and search*, volume 81. The MIT Press, 2000. URL `/home/conormw/Dropbox/Articles/Philosophy/Analytic/LinktoCarnap-Meaning_Postulates.pdf`.

Michael Strevens. The role of the priority rule in science. *The Journal of philosophy*, 100(2):55—79, 2003.

Michael Strevens. The role of the matthew effect in science. *Studies In History and Philosophy of Science Part A*, 37(2):159—170, 2006.

Michael Weisberg and Ryan Muldoon. Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2):225—252, April 2009.

Uri Wilensky. NetLogo. *Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.*, 1999. URL `http://ccl.northwestern.edu/netlogo/`.

H. Peyton Young. The dynamics of social innovation. *Proceedings of the National Academy of Sciences, forthcoming*, 2011.

Kevin J. Zollman. Social structure and the effects of conformity. *Synthese*, 172(3):317—340, 2010.

Kevin J. Zollman. The communication structure of epistemic communities. In A Goldman, editor, *Social Epistemology: Essential Readings*, pages 338—350. 2011a.

Kevin J. Zollman. Systems-Oriented social epistemology and a humean approach to testimony, 2011b.

5 Appendix A - Proofs

5.1 Notation

Given a set S, let $\mathcal{P}(S)$ denote its power set. Let S^T denote all functions from T to S, and so $S^{\mathbb{N}}$ is the set of all functions from natural numbers to S. Alternatively, one can think of $S^{\mathbb{N}}$ as all infinite sequences of elements of S. So we suggestively define $S^{<\mathbb{N}}$ to be all finite sequences from S.

For any set S, let $|S|$ denotes its cardinality; when S is a sequence, $|S|$ is therefore its length. Given a sequence σ and $n \leq |\sigma|$, let σ_n denote the n^{th} coordinate of σ. If the coordinates of σ are likewise sequences, then let $\sigma_{n,k}$ be the k^{th} coordinate of the n^{th} coordinate of σ. And so on if σ is a sequence of sequences of sequences. Alternatively, let π_n denote the n^{th} projection function so that $\sigma_n = \pi_n(\sigma)$, and $\sigma_{n,k} = \pi_k(\pi_n(\sigma))$. Let $\sigma \restriction n$ denote the initial segment of σ of length n.

If $S_1, S_2, \ldots, S_n$ is a sequence of sets, then let $\times_{j \leq n} S_j$ be the standard Cartesian product. Given a collection of σ-algebras $\langle S_i, \mathcal{S}_i \rangle_{i \in I}$, let $\otimes_{i \in I} \mathcal{S}_i$ denote the product algebra. In particular, $\otimes_{n \in \mathbb{N}} \mathcal{S}$ is the infinite product space generated by a single σ-algebra.

Give a σ-algebra $\mathcal{S}$, let $\mathbb{P}(\mathcal{S})$ denote the set of all probability measures on $\mathcal{S}$. If p is a probability measure on $\langle S, \mathcal{S} \rangle$, then let $p^n \in \mathbb{P}(\otimes_{k \leq n} \mathcal{S})$ denote the induced product measure on $\otimes_{k \leq n} \mathcal{S}$. When $\mathcal{S}$ is a Borel algebra, these measures extend uniquely to a measure p^∞ on $\otimes_{n \in \mathbb{N}} \mathcal{S}$ (i.e., p^∞ is the unique measure on $\langle S^{\mathbb{N}}, \otimes_{n \in \mathbb{N}} \mathcal{S} \rangle$ such that $p^\infty(F_1 \times F_2 \ldots \times F_n \times S^{\mathbb{N}}) = p(F_1) \cdot p(F_2) \cdots p(F_n)$, where $F_i \in \mathcal{S}$ for all $i \leq n$). Given a metric space M, let $\mathbb{B}(M)$ denote the Borel algebra. Given $\epsilon > 0$ and $m \in M$, let $B_\epsilon(m)$ denote the ϵ ball around m.

5.2 Preliminaries

This section introduces basic facts about Markov processes that will be used in later proofs. For proofs, see any introductory exposition of Markov processes, such as Chapter 11 in Grinstead and Snell [1997].

A sequence of (abstract) random variables $\langle X_n \rangle_{n \in \mathbb{N}}$ is called **Markov process** just in case $p(X_n | X_1, X_2 \ldots X_n) = p(X_{n+1} | X_n)$ for all n. Given a Markov process, the codomain S of the random variables is called the **state space**, and its elements are called **states.** For the remainder of this paper, I will discuss only Markov processes that have finitely many states. A Markov chain is called **ergodic** if, for any pair of states s_i and s_j, there is some finite number n of steps for which it is possible to for the process to transition from s_i to s_j in n steps. More formally, for any pair of states s_i

and s_j, there is some finite number n such that $p(X_{m+n} = s_j | X_m = s_i) > 0$ for all m.

A Markov process is **time-homogeous** if the probability of transitioning from one state to the next is the same at all times. Formally, this means that $p(X_{n+1} = s_i | X_n = s_j) = p(X_{m+1} = s_i | X_m = s_j)$ for all $s_i, s_j \in S$ and $m, n \in \mathbb{N}$. Since S is finite, one can represent a time-homogenous Markov process by a transition matrix $\boldsymbol{P}$ whose ij^{th} entry is the probability $p(X_{n+1} = s_i | X_n = s_j)$. Let

$$\boldsymbol{P}^n = \underbrace{\boldsymbol{P} \cdot \boldsymbol{P} \cdots \boldsymbol{P}}_{\text{n-times}}$$

It is easy to show that the ij^{th} entry of the matrix $\boldsymbol{P}^n$ the probability of the problem being in state s_i exactly n stages after it is in state s_j. For the remainder of the Appendix, we will consider only time-homogeneous Markov processes.

If $\boldsymbol{P}^m$ is a strictly positive matrix for some $m \in \mathbb{N}$, then the Markov process is called **regular**. Informally, a Markov process is regular if it is possible to transition from any state to any other in exactly m steps. The following theorems concerning Markov processes are well-known, and so their proofs are omitted.

Theorem 8 *If $\langle X_n \rangle_{n \in \mathbb{N}}$ is a finite, regular, time homogenous Markov process with transition matrix $\boldsymbol{P}$, then there exists some transition matrix $\boldsymbol{P}_\infty$ such that $\lim_{n \to \infty} \boldsymbol{P}^n = \boldsymbol{P}_\infty$. Moreover, all rows of $\boldsymbol{P}_\infty$ are identical (i.e., columns are constant).*

Theorem 9 *Let $\boldsymbol{P}$ be a transition matrix for a time-homogeneous, ergodic Markov process. If the diagonal entries of $\boldsymbol{P}$ are all positive, then $\boldsymbol{P}$ is regular.*

A set of states $S_* \subseteq S$ is said to be **absorbing** if $p(X_{n+1} \in S_* | X_n \in S_*) = 1$. A Markov process is **absorbing** if it is ergodic and there is at least one absorbing state s_*.

Theorem 10 *Suppose $\langle X_n \rangle_{n \in \mathbb{N}}$ is an absorbing Markov process with absorbing states S_*. Then $p(\lim_{n \to \infty} X_n \in S_*) = 1$.*

In fact, the above theorem can be strengthened slightly. One need not assume the Markov process is ergodic, but only that, for each state s, there is an absorbing state $s_* \in S_*$ (which may depend upon s) such that there is positive probability of transiting from s to s_*. This is the stronger version of the theorem used in the remainder of the Appendix.

5.3 Piecewise Conditional Markov Processes

Consider a sequence $\langle X_n, E_n \rangle_{n \in \mathbb{N}}$, where the X_n's are random variables and the E_n's are events. The sequence is called a **piecewise conditional Markov process** (or pc-Markov process) if

1. The events $\langle E_n \rangle_{n \in \mathbb{N}}$ are pairwise disjoint,
2. $p(\cup_{n \in \mathbb{N}} E_n) = 1$, and
3. $\langle X_k \rangle_{k \geq n}$ is a Markov process with respect $p(\cdot | E_n)$, i.e., for all $k \geq n$:

$$p(X_{k+1} | E_n, X_1, X_2, \ldots X_k) = p(X_{k+1} | E_n, X_k)$$

Call the sequences of random variables $\langle X_k \rangle_{k \geq n}$ the **pieces** of a pc-Markov process. Say a pc-Markov process is ergodic/regular/absorbing if each of its pieces is ergodic/regular/absorbing. Say it is **uniformly regular** if the pieces are regular and have identical transition matrices, and similarly, say it is **uniformly absorbing** if each of the pieces are absorbing and have an identical set of absorbing states. The next two theorems show that the asymptotic behavior of uniformly absorbing (or regular) pc Markov processes is identical to that of each of their pieces.

Theorem 11 *Suppose $\langle X_n, E_n \rangle_{n \in \mathbb{N}}$ is a uniformly absorbing, pc-Markov process with absorbing states S_*. Then $p(\lim_{n \to \infty} X_n \in S_*) = 1$.*

Proof:

$$\begin{aligned}
p(\lim_{n \to \infty} X_n \in S_*) &= p(\cup_{n \in \mathbb{N}} E_n \cap (\lim_{n \to \infty} X_n \in S_*)) \\
&= p(\cup_{n \in \mathbb{N}} (E_n \cap \lim_{k \to \infty} X_k \in S_*)) \\
&= \sum_{n \in \mathbb{N}} p(E_n \cap \lim_{k \to \infty} X_k \in S_*) \\
&= \sum_{n \in \mathbb{N}} p(\lim_{k \to \infty} X_k \in S_* | E_n) \cdot p(E_n) \\
&= \sum_{n \in \mathbb{N}} p(\lim_{k \geq n} X_k \in S_* | E_n) \cdot p(E_n) \\
&= \sum_{n \in \mathbb{N}} 1 \cdot p(E_n) \\
&= 1
\end{aligned}$$

The first equality follows by the fact that $p(\cup_{n \in \mathbb{N}} E_n) = 1$; the second follows from Demorgan's laws. The third is a consequence of the pairwise disjointness of the E_n's, and the fourth follows by the definition of conditional probability. The fifth follows from basic facts about limits, and the sixth follows from Theorem 10 and the fact that $\langle X_k \rangle_{k \geq n}$ is an absorbing Markov chain under $p(\cdot | E_n)$. Finally, the last equality follows from the fact that $p(\cup_{n \in \mathbb{N}} E_n) = 1$ and the disjointness of the E_n's.

□

Theorem 12 *Suppose $\langle X_n, E_n \rangle_{n \in \mathbb{N}}$ is a uniformly regular, pc-Markov process with transition matrix $\boldsymbol{P}$. Then for any state $s_i \in S$ there is some probability $r_i \in [0,1]$ such that $\lim_{n \to \infty} p(X_n = s_i) = r_i$.*

Proof: Let $\boldsymbol{P}$ be the common transition matrix of the uniformly regular, pc-Markoc process. By Theorem 8, there is some $\boldsymbol{P}_\infty$ with constant columns such that $\boldsymbol{P}^n \to \boldsymbol{P}_\infty$ as $n \to \infty$. Given a state s_i, let r_i be the value of the i^{th} column in $\boldsymbol{P}_\infty$. So r_i is the asymptotic probability of state s_i for a Markov process with transition matrix $\boldsymbol{P}$. We claim that $\lim_{n \to \infty} p(X_n = s_i) = r_i$.

To prove this, we must find some n_ϵ such that $|p(X_k = s_i) - r_i| < \epsilon$ for all $k \geq n_\epsilon$. First, note that because $\sum_{n \in \mathbb{N}} p(E_n) = 1$, there is some k_ϵ such that $\sum_{n \leq k_\epsilon} p(E_n) \geq 1 - \frac{\epsilon}{2}$. As the pieces are regular Markov processes with transition matrix $\boldsymbol{P}$, it follows that $\lim_{k \to \infty} p(X_k = s_i | E_n) = r_i$ for all n by theorem 8. Hence, for all n, there is some $m_{\epsilon,n}$ such that

$$|p(X_k = s_i | E_n) - r_i| < \frac{\epsilon}{2}$$

for all $k \geq m_{\epsilon,n}$. Let $n_\epsilon = \max\{m_{\epsilon,n} : m \leq k_\epsilon\}$. Then for all $k \geq n_\epsilon$, it follows that:

$$\begin{array}{rcl} p(X_k = s_i) & = & p(X_k = s_i \cap \cup_{n \in \mathbb{N}} E_n) \\ & = & p(\cup_{n \in \mathbb{N}}(X_k = s_i \cap E_n)) \\ & = & \displaystyle\sum_{n \in \mathbb{N}} p(X_k = s_i \cap E_n) \\ & = & \displaystyle\sum_{n \in \mathbb{N}} p(E_n) \cdot p(X_k = s_i | E_n) \\ & \leq & \displaystyle\frac{\epsilon}{2} + \sum_{n \leq k_\epsilon} p(E_n) \cdot p(X_k = s_i | E_n) \text{ by choice of } k_\epsilon \\ & \leq & \displaystyle\frac{\epsilon}{2} + \sum_{n \leq k_\epsilon} p(E_n) \cdot (r_i + \frac{\epsilon}{2}) \text{ by choice of } n_\epsilon \text{ and the fact that } k \geq n_\epsilon \\ & = & \displaystyle\frac{\epsilon}{2} + (r_i + \frac{\epsilon}{2}) \sum_{n \leq k_\epsilon} p(E_n) \\ & \leq & \displaystyle\frac{\epsilon}{2} + (r_i + \frac{\epsilon}{2}) \\ & = & r_i + \epsilon \end{array}$$

Similarly,

$$\begin{array}{rcl} p(X_k = s_i) & \geq & \displaystyle\sum_{n \leq k_\epsilon} p(E_n) \cdot p(X_k = s_i | E_n) \\ & \geq & \displaystyle\sum_{n \leq k_\epsilon} p(E_n) \cdot (r_i - \frac{\epsilon}{2}) \\ & = & \displaystyle(r_i - \frac{\epsilon}{2}) \sum_{n \leq k_\epsilon} p(E_n) \\ & \geq & \displaystyle r_i - \frac{\epsilon}{2} \end{array}$$

So $|p(X_k = s_i) - r_i| < \epsilon$ as desired.

□

5.4 Definitions

5.4.1 Worlds and Questions

Define a **question** to be a triple $\langle W, \langle \Theta, \rho \rangle \rangle$, where W is an arbitrary set whose elements are called **worlds**, Θ is a partition of W, and ρ is a metric on Θ. Elements of Θ are called **answers** to the question, and ρ quantifies how different two possible answers are. I use the variable Θ because, as will be seen in examples below, in many circumstances Θ is a partition of possible parameter values for some unknown distribution, and the variable Θ is frequently used in statistics to represent parameter values. Given a world w, let $\theta_w \in \Theta$ be the partition cell (i.e. answer) containing w.

Example 1: Suppose one is interested in the bias of a trick coin (i.e., the frequency with which the coin lands heads). In this case, the set of possible worlds W is the set of real numbers between 0 and 1 inclusive (i.e. $W = [0, 1]$), and the set of possible answers is $\Theta := \{\{\theta\} : \theta \in [0, 1] = W\}$. The appropriate metric ρ on Θ is the Euclidean metric: $\rho(\{\theta_1\}, \{\theta_2\}) = |\theta_1 - \theta_2|$. Let Q_B be the question described here; B stands for "Bernoulli."

Example 2: Suppose one is interested in the mean of a normal distribution. However, imagine one does not care about the exact value of the mean, but rather one only wants to know whether it is at least zero. In this case, the set of possible worlds W is the set of ordered pairs $\langle \mu, \sigma^2 \rangle \in \mathbb{R} \times \mathbb{R}^+$ representing the mean and variance of the unknown distribution. The set of possible answers is $\Theta := \{\theta_{\geq 0}, \theta_{<0}\}$, where $\theta_{\geq 0} = \{\langle \mu, \sigma^2 \rangle \in W : \mu \geq 0\}$ and $\theta_{<0} = \{\langle \mu, \sigma^2 \rangle \in W : \mu < 0\}$.

Here, the most appropriate metric ρ on Θ would be the discrete metric, which assigns a distance of 1 between any two distinct answers. Let Q_N be the question described here; here, N stands for "normal." This is the question described in depth in the running example in the body of the paper.

Example 3: This example presupposes some familiarity with the practice of representing causal relationships among random variables by means of Bayesian networks. Suppose one is interested in the causal relationships among a collection of real-valued, random variables $\mathcal{V} = \{V_1, \ldots, V_k\}$. Let $\langle \Omega, \mathcal{F} \rangle$ be the the measurable space on which the V_j's are defined. Suppose, however, that one has access only to observational data, and hence, one believes can discover causal structure only up to Markov equivalence [Spirtes et al., 2000].

The set of worlds W consists of all pairs $\langle G, p \rangle$ such that $p \in \mathbb{P}(\mathcal{F})$ is Markov and faithful to the graph $G \in \textsc{dag}_{\mathcal{V}}$, where $\textsc{dag}_{\mathcal{V}}$ is the set of directed acyclic graphs with vertex set $\mathcal{V}$. Given two worlds $w = \langle G, p \rangle$ and $w' = \langle G', p' \rangle$, write $w \equiv w'$ if G is Markov equivalent to G'. Let Θ be the partition on W induced by equivalence relation $\equiv$. If one is only interested in discovering the underlying causal structure/graph (up to Markov equivalence), then the discrete metric is the appropriate measure of distance between two elements of Θ. Call this question Q_C, where C stands for "causal graph."

Example 4: Suppose one's data is drawn from a Markov process that is known to have a unique absorbing state, and one is interested in discovering that state. So let S be a finite state space, and W be the set of transition matrices for absorbing Markov processes over S with a unique absorbing state. Define the distance between two such transition matrices as follows:

$$\rho(w, w') = \sqrt{\sum_{s_i \in S} \sum_{s_j \in S} |w_{ij} - w'_{ij}|^2}.$$

Call this question Q_M, where M stands for "Markov."

5.4.2 Learning Problems

Imagine that one learns about the world by making a series of observations. To model such learning, define a **data generating process** for a question $Q = \langle W, \langle \Theta, \rho \rangle \rangle$ to be a pair $\langle \langle D, \mathcal{D} \rangle, c \rangle$ where

1. $\langle D, \mathcal{D} \rangle$ is a measurable space, and
2. $c : W \to \mathbb{P}(\otimes_{n \in \mathbb{N}} \mathcal{D})$ is a function, whose values c_w are called the **chances** under w.

Define a **learning problem** L to be a pair consisting of a question and a data generating process for that question. Informally, the set D represents **data** that one might learn when making observations in any world. For any world w, the probability measure c_w specifies how likely one is to observe particular data sequences. These notions are best explained by examples:

Example 1: Consider $Q = Q_B$ be the Bernoulli question in Example 1 above. One data-generating process for Q_B is the following. Let $D = \{0, 1\}$ and let $\mathcal{D} = \mathcal{P}(\{0, 1\})$ be the power set algebra. Here, 0 represents observing a heads, and 1 represents observing a tails. For all $w \in W = [0, 1]$, let p_w be the unique measure on $\mathcal{D}$ such that $p_w(\{0\}) = w$. Define $c_w = (p_w)^\infty$. Then the chances given by c_w are exactly those of independent coin tosses.

Example 2: Let $Q = Q_N$ be the "normal" question in Example 2 above. Let $D = \mathbb{R}$ and let $\mathcal{D} = \mathbb{B}(\mathbb{R})$ be the Borel algebra. So data are just sample points pulled from a normal distribution. For every world $w = \langle \mu, \sigma \rangle \in$

$\mathbb{R} \times \mathbb{R}^+$, let p_w be the unique measure on $\mathcal{D}$ such that the density of p_w is a normal distribution with mean μ and variance σ^2. Let $c_w = (p_w)^\infty$, and let L_N be learning problem described here.

Example 3: Let $Q = Q_C$ be the causal question in Example 3 above. Let $D = \mathbb{R}^{|\mathcal{V}|}$ and $\mathcal{D} = \otimes_{k \leq |\mathcal{V}|}\mathbb{B}(\mathbb{R})$. For all $w = \langle G, p\rangle \in W$, let p_w be the unique measure on $\mathcal{D}$ such that $p_w(E) = p(\langle V_1, V_2 \ldots V_{|\mathcal{V}|}\rangle \in E)$ for $E \in \mathcal{D}$. Again, let $c_w = (p_w)^\infty$, and let L_C be learning problem described here.

Example 4: In the previous three examples, the data generated by the underlying world consists of the values of iid random variables. One need not assume that data is acquired from such a process. For instance, let $Q = Q_M$ be the "Markov" question in Example 4 above, and suppose one observes successive states of the underlying Markov process.

To model such learning, let $D = S$ be the states of the Markov process and $\mathcal{D}$ be the power set of D. Fix a state $s_0 \in S = D$, which represents the starting state of the process. For all transition matrices $w \in W$, let c_w be the unique product measure on the infinite product space on $S^{\mathbb{N}}$ such that

$$c_w(\langle s_1, s_2, \ldots, s_n\rangle \times S^{\mathbb{N}}) = \Pi_{k \leq n} w_{s_i, s_{i+1}} = w_{s_0, s_1} \cdot w_{s_1, w_2} \cdots w_{s_{n-1}, s_n}$$

where w_{s_i, s_j} is the element of the transition matrix w giving the probability of transitioning from s_i to s_j. Then, in general, the data generated in this process are highly dependent and not identically distributed. Let L_M be the learning problem defined here.

5.5 Methods

A **method** for a learning problem L is a function $m : D^{<\mathbb{N}} \to \mathbb{P}(\mathbb{B}(\Theta))$. Let $m_d := m(d)$ for all $d \in D^{<\mathbb{N}}$. Informally, a method takes data sequences as input and returns sets of answers with different probabilities.

Example 1: Consider the learning problem L_B. Then one common method m is return the sample mean. In other words, let $d \in \{0, 1\}^n$ be a sequence of coin flips (e.g. $\langle 0, 1, 1\rangle$) of some finite length n , and let $w(d)$ be the proportion of flips in d that land heads (e.g. $w(\langle 0, 1, 1\rangle) = \frac{2}{3}$). Notice that since the possible set of worlds W is the unit interval $[0, 1]$, the proportion $w(d) \in W$. Then define m to be the function such that, for all d, the measure m_d assigns unit probability to the event $\{w(d)\}$ and zero probability elsewhere.

Example 2: Consider the learning problem L_N. One method m is to employ a likelihood ratio test to test the null hypothesis $H_0 : \mu \geq 0$ versus the alternative, where the significance of the test is decreased at a rate of the natural log of the sample size. Formally, let $d \in \mathbb{R}^n$ be a data sequence, and let $\mu(d)$ and $\sigma^2(d)$ be the (sample) mean and variance of the data d. Let p_d be the probability measure on $\mathbb{R}$ such that the density of p_d is a normal distribution with mean 0 and variance $\sigma^2(d)$. Let $\alpha \in (0, 1)$ be a fixed significance level, and define a method m such that m_d assigns (i) probability one to $\theta_{\mu \geq 0}$ if $p_d(x \in \mathbb{R} : x \geq \mu(d)\}) \geq \frac{1-\alpha}{\ln |d|}$ and (ii) probability one to $\theta_{\mu<0}$ otherwise.

Example 3: Consider the learning problem $L_{\mathcal{V}}$. Spirtes et al. [2000] define several methods including the SGS algorithm, PC algorithm, and so on.

Example 4: Consider the learning probelm L_M. Define a method m that returns the last observed state of the process. In other words, let $d = \langle s_1, s_2 \ldots, s_n\rangle$ be a data sequence of observed states. Then m_d assigns probability one to the event $\{s_n\}$.

Three comments about methods are in order. First, as indicated in the above samples, the most commonly employed methods in statistics are **deterministic** in the sense that for all data sequences d, there exists some $\theta \in \Theta$ such that $m_d(\theta) = 1$. In other words, answers are chosen deterministically in response to data sequences and do not depend upon some external randomizing device. Although none of the examples I have described require indeterministic methods, I have defined methods in such a way to accomodate such generality.

Second, when the learning problem L concerns discovering some unknown parameter, then a deterministic method in my sense is simply a parameter estimator in the standard statistical sense (see Example 1 above).

Finally, in statistics, it is standard to define methods to be functions from *data* to parameters, or decisions, etc. Since my methods return answers with different probabilities, it is easier to think of methods as returning *probability measures* over answers; to determine how much probability is assigned to different sets of answers given some data d, we investigate the measure m_d that is *indexed* by d.

5.6 Convergence

Methods are often evaluated by whether they discover the true answer in a given world, and how quickly they do so. To this end, we introduce several notions of convergence, some of which are standard in probability theory in statistics. To do so, however, we must first define a probability measure over infinite sequences of answers so that we can characterize the asymptotic performance of a method. Fix some natural number n. Given a method m and world w, define $p^n_{w,m}$ to be the unique measure on $\langle \Theta^n, \otimes_{k \leq n} \mathbb{B}(\Theta) \rangle$ satisfying the following. For all "rectangles" $E_1 \times E_2 \times \ldots \times E_n \in \otimes_{k \leq n} \mathbb{B}(\Theta)$ (i.e., $E_k \in \mathbb{B}(\Theta)$ for all $k \leq n$):

$$p^n_{w,m}(E_1 \times E_2 \times \ldots \times E_n) = \int_{D^n} \prod_{k \leq n} m_{\delta \restriction k}(E_k) \; dc^n_w(\delta)$$

where (1) c^n_w is the restriction of c_w to $\langle D^n, \otimes_{k \leq n} \mathcal{D} \rangle$ (i.e. c^n_w is the unique measure such that $c^n_w(F) = c_w(F \times D^{\mathbb{N}})$ for all $F \in \otimes_{k \leq n} \mathcal{D}$), and (2) δ is an element of D^n (i.e., a data sequence of length n). I have used δ instead of d to denote a data sequence so as to avoid confusion because the dummy letter "d" is used here to indicate that the integral is being taken with respect to the measure c^n_w. Under the measure $p^n_{w,m}$, the probability of a method returning a sequence of answers is the probability of obtaining particular data sequence (given by c^n_w) times the probability that the method returns a given answer (given by m) in response to that data sequence. So this construction assumes that the randomizing device by which one chooses answers is probabilistically independent of the data that will be observed in a world.

Using standard measure-theoretic constructions, it is easy to show that there is a unique probability measure $p_{w,m}$ on the infinite product algebra $\langle \Theta^{\mathbb{N}}, \otimes_{n \in \mathbb{N}} \mathbb{B}(\Theta) \rangle$ extending each of the $p^n_{w,m}$'s (i.e. $p_{w,m}(E \times \Theta^{\mathbb{N}}) = p^n_{w,m}(E)$ for all $E \in \otimes_{k \leq n} \mathbb{B}(\Theta)$ and all natural numbers n). Further, it is easy to check the following are events in the infinite product algebra, where $\theta \in \Theta$ and $\epsilon > 0$:

$$\begin{aligned} \{\overline{\theta}_n = \theta \text{ for large n}\} &:= \{\overline{\theta} \in \Theta^{\mathbb{N}} : (\exists n \in \mathbb{N})(\forall k \geq n)\overline{\theta}_k = \theta\} \\ \{\lim_{n \to \infty} \overline{\theta}_n = \theta\} &:= \{\overline{\theta} \in \Theta^{\mathbb{N}} : \lim_{n \to \infty} \overline{\theta}_n = \theta\} \\ \{\overline{\theta}_n \in B_\epsilon(\theta)\} &:= \{\overline{\theta} \in \Theta^{\mathbb{N}} : \overline{\theta}_n \in B_\epsilon(\theta)\}. \end{aligned}$$

A method m is called

- **almost surely** (a.s.) convergent if for all $w \in W$

$$p_{w,m}(\lim_{n\to\infty} \overline{\theta}_n = \theta_w) = 1$$

- **consistent** if for all $w \in W$ and all $\epsilon > 0$

$$\lim_{n\to\infty} p_{w,m}(\overline{\theta}_n \in B_\epsilon(\theta_w)) = 1$$

- **uniformly consistent** if for all $\epsilon > 0$

$$\lim_{n\to\infty} \inf_{w\in W} p_{w,m}(\overline{\theta}_n \in B_\epsilon(\theta_w)) = 1$$

- **strongly almost surely** (s.a.s.) convergent if for all $w \in W$

$$p_{w,m}(\overline{\theta}_n = \theta_w \text{ for large } n) = 1$$

When $\Theta = \{\{r\} : r \in \mathbb{R}^d\}$ is some parametric model, then almost sure convergence (respectively, convergence and uniform convergence) of a method is the standard notion of almost sure convergence (respectively, convergence and uniform convergence) of a parameter estimator.

The relationships between the notions of convergence are as follows; the first two facts are well-known and the last is trivial.

Lemma 1 *Let $\widehat{\theta}$ be a method for learning problem L.*

1. *If m converges a.s., then it is consistent, but not vice versa.*
2. *If m is uniformly consistent, then it is consistent, but not vice versa.*
3. *Suppose Θ is finite. Then a.s. convergence entails s.a.s. convergence.*

Example 1: Consider the learning problem L_B. Then the method that returns the sample mean (i.e. frequency of heads) converges almost surely by the strong law of large numbers. Hence, it is consistent as well. However, it is not s.a.s., as if w is an irrational number, then the method never returns w with any positive probability at any sample size.

Example 2: Consider the learning problem L_N. The method defined above is consistent. See Jeffreys [1998]

Example 3: Consider the learning probelm $L_{\mathcal{V}}$. Then the algorithms of Spirtes et al. [2000] are all consistent.

Example 4: Consider the learning probelm L_M. Then the method of returning the last state describe above is a.s. convergent, and hence, s.a.s. convergent because the process is finite. Why? By Theorem 10, with probability one, the process eventually enters the unique absorbing state s_* and never leaves it. Since the method m returns that last observed state with probability one (under the measure m_d), then there is unit probability (under $p_{w,m}$) that m eventually returns the unique absorbing state from some stage onward. The same argument works for a method that returns the k^{th} to last observed state, for some fixed natural number k.

5.6.1 Expert Networks

A **network** is a finite undirected graph G; we will refer to the vertices of G as **agents**. A **group** is a set of agents $J \subseteq G$. For any $g \in G$, let $N_G(g) \subseteq G$ denote the group of agents $g' \in G$ such that g and g' are incident to a common edge. Call $N_G(g)$ the **neighborhood** of g, and call is elements **neighbors** of g. For simplicity, we assume every agent is her own neighbor. When the network is clear from context, I will write $N(g)$ instead of $N_G(g)$.

An **expert network** $\mathcal{E}$ is a pair $\langle G, \langle L_g \rangle_{g \in G}, \langle m_g \rangle_{g \in G} \rangle$ such that G is a network, L_g is a learning problem for each agent $g \in G$, and m_g is a method for L_g. For each agent $g \in G$, let Q_g be the question confronted by the agent; define Θ_g, $c_{w,g}$, etc., similarly. An expert network can be represented by a colored undirected graph such that the vertices g and g' are the same color just in case $Q_g = Q_{g'}$; notice, agents may share a color in the graph even if the type of data they collect is very different. In other words, agents that are faced with the same question might observe data that cannot be combined in any meaningful way.

Example: In the running example in the body of the paper, the expert networks consist of agents confronted with instances of learning problem L_N; however, different agents may sample from different normal distributions.

Let $\Theta_{\mathcal{E}} = \{\Theta_g : g \in G\}$ be the set of questions faced by agents in the expert network $\mathcal{E}$, and let $A_{\mathcal{E}} = \times_{\Theta \in \Theta_{\mathcal{E}}} \Theta$ be the set of answers to all questions raised in the expert network. Define $\Theta_{\mathcal{E}-g} = \Theta_{\mathcal{E}} \setminus \{\Theta_g\}$ to be the set of questions faced by agents other than g, and $A_{\mathcal{E}-g} = \times_{\Theta \in \Theta_{\mathcal{E}-g}} \Theta$ be all possible answers.

In future sections, we will need to investigate several agents' beliefs (i) to all questions under investigation in a network (ii) over several stages of inquiry. In symbols, if $J \subseteq G$ is a group, then a sequence of answers given by J to all questions under investigation is a member of the set $((A_{\mathcal{E}})^J)^{<\mathbb{N}}$. We realize this is a notational nightmare because the elements of these sets are sequences of sequences of sequences. However, the complicated notation is unavoidable. To avoid confusion, we will use specific letters to denote elements of particular sets. The letter "θ" will always be an answer to a *single* question Θ. The letter "a" will designate answers to *several* questions, and hence, will be a sequence of answers; generally, a will be a member of $A_{\mathcal{E}}$ or of $A_{\mathcal{E}-g}$. The bolded letter $\boldsymbol{a}$ will indicate several agents' (i.e. a group's) answers to several questions (so $\boldsymbol{a} \in (A_{\mathcal{E}})^J$ or $\boldsymbol{a} \in (A_{\mathcal{E}-g})^J$). Finally, we use the "bar-notation" $\overline{\boldsymbol{a}}$ to indicate a sequence of group answers to several questions (so $\overline{\boldsymbol{a}} \in ((A_{\mathcal{E}})^J)^{<\mathbb{N}}$).

Recall, by definition of a question, there is a Borel algebra $\mathbb{B}(\Theta)$ over each $\Theta \in \Theta_{\mathcal{E}}$. Hence, one can define $\mathcal{A}_{\mathcal{E}}$ be the product σ-algebra on $A_{\mathcal{E}} = \times_{\Theta \in \Theta_{\mathcal{E}}} \Theta$, and similarly for $\mathcal{A}_{\mathcal{E}-g}$. It is easy to check the following are events in these algebras:

$$\begin{aligned} \{(\forall g \in G) \overline{\boldsymbol{a}}_{n,g} = a \text{ for large } n\} &= \{\overline{\boldsymbol{a}} \in ((A_{\mathcal{E}})^G)^{\mathbb{N}} : (\exists n \in \mathbb{N})(\forall k \geq n)(\forall g) \overline{\boldsymbol{a}}_{n,g} = a\} \\ \{(\forall g \in G) \lim_{n \to \infty} \overline{\boldsymbol{a}}_{n,g} = a\} &= \{\overline{\boldsymbol{a}} \in ((A_{\mathcal{E}})^G)^{\mathbb{N}} : (\forall g \in G) \lim_{n \to \infty} \overline{\boldsymbol{a}}_{n,g} = a\} \end{aligned}$$

And so on. These definitions will be of importance in the definition of a testimonial norm, and in characterizing their asymptotic reliability.

5.6.2 Testimonial Norms

A **testimonial norm** is a class of functions $\tau_{\mathcal{E},g} : ((A_{\mathcal{E}-g})^{N(g)})^{<\mathbb{N}} \to \mathbb{P}(\mathcal{A}_{\mathcal{E}-g})$, where the index $\mathcal{E}$ ranges over expert networks and g ranges over agents in the network $\mathcal{E}$. Informally, a testimonial norms specifies, for each expert network $\mathcal{E}$ and each agent $g \in \mathcal{E}$, a probability distribution over answers to questions outside an agent g's area of expertise given what g's neighbors have reported in the past.

The body of the paper introduces the following six testimonial norms. To describe them, let $\overline{\boldsymbol{a}} \in ((A_{\mathcal{E}-g})^{N(g)})^{<\mathbb{N}}$ be an arbitrary sequence of answer reports of g's neighbors to all questions of interest. Suppose $\overline{\boldsymbol{a}}$ has length n. Recall that, by our notational conventions, for each agent h in g's neighborhood and each $\Theta \in \Theta_{\mathcal{E}-g}$, the symbol $\overline{\boldsymbol{a}}_{n,h,\Theta}$ represents h's report to question Θ on stage n (i.e. the last stage of a). Similarly, if $a \in A_{\mathcal{E}-g}$ is an answer to all questions outside of g's area of expertise, and if $\Theta \in \Theta_{\mathcal{E}-g}$ is one such question outside of g's area of expertise, then a_Θ is the answer a provides to the question Θ.

Example 1: Reidianism is the norm such that for all $a \in A_{\mathcal{E}-g}$:

$$\tau_{\mathcal{E},g}(\overline{\boldsymbol{a}})(a) = \prod_{\Theta \in \Theta_{\mathcal{E}-g}} \frac{|\{h \in N(g) : a_\Theta = \overline{\boldsymbol{a}}_{n,h,\Theta}\}|}{|N(g)|}$$

In other words, an answer θ to a given question Θ is chosen to be the proportion of one's neighbors that report θ on the most recent stage. Answers to different questions are chosen independently of one another, so the probability of choosing a sequence of answers a is the product of the probabilities of choosing each element a_Θ of the sequence.

To describe the remaining norms, we need to introduce some notation. Let $N_{\mathcal{E}}(g,\Theta) = \{h \in N_G(g) : \Theta_h = \Theta\}$ be the neighbors of g who study the question Θ; if there are no such neighbors, we stipulate that $N_{\mathcal{E}}(g,\Theta) = N(g)$. Let $PN_{\mathcal{E}}(g,\Theta)$ be those neighbors of g that have minimal path length to an expert in Θ. Again, when $\mathcal{E}$ is clear from context, I will drop the subscript.

Example 2: E-trusting is the norm such that for all $a \in A_{\mathcal{E}-g}$:

$$\tau_{\mathcal{E},g}(\overline{\boldsymbol{a}})(a) = \prod_{\Theta \in \Theta_{\mathcal{E}-g}} \frac{|\{h \in N(g,\Theta) : a_\Theta = \overline{\boldsymbol{a}}_{n,h,\Theta}\}|}{|N(g)|}$$

Example 3: Proxmitism is the norm such that for all $a \in A_{\mathcal{E}-g}$:

$$\tau_{\mathcal{E},g}(\overline{\boldsymbol{a}})(a) = \prod_{\Theta \in \Theta_{\mathcal{E}-g}} \frac{|\{h \in PN(g,\Theta) : a_\Theta = \overline{\boldsymbol{a}}_{n,h,\Theta}\}|}{|N(g)|}$$

Although it is a tedious task, the majoritarian versions of all three norms can also be defined rigorously in a similar manner. We leave this task to the reader.

As noted in the body of the paper, there are a few special properties of testimonial norms that will play a critical role in proofs. Let τ be a testimonial norm. Suppose that for all expert networks $\mathcal{E}$, all agents g in $\mathcal{E}$, and all answer reports $\overline{\boldsymbol{a}}, \overline{\boldsymbol{b}} \in ((A_{\mathcal{E}-g})^{N(g)})^{<\mathbb{N}}$ with identical last coordinates (i.e., $\overline{\boldsymbol{a}}_{|\overline{\boldsymbol{a}}|} = \overline{\boldsymbol{b}}_{|\overline{\boldsymbol{b}}|}$):

$$(*)\ \tau_{\mathcal{E},g}(\overline{\boldsymbol{a}}) = \tau_{\mathcal{E},g}(\overline{\boldsymbol{b}}).$$

Then τ is said to be **time homogeneous** and **Markov**, as its behavior depends only upon the last element of an answer sequence (time-homogeneity is implicitly built in as $\overline{\boldsymbol{a}}$ and $\overline{\boldsymbol{b}}$ need not even have the same length). It is said to be time-homogeneous and Markov with memory t if $(*)$ holds for any sequences $\overline{\boldsymbol{a}}$ and $\overline{\boldsymbol{b}}$ for which the last t coordinates are identical.

Given $\Theta \in \Theta_{\mathcal{E}-g}$ and some $\theta \in \Theta$, define $E(\theta) = \{a \in A_{\mathcal{E}-g} : a_\Theta = \theta\}$. A time-homogeneous, Markov testimonial norm τ with memory t is said to be **stable** if, for all $\overline{\boldsymbol{a}} \in ((A_{\mathcal{E}-g})^{N(g)})^{<\mathbb{N}}$, if $\overline{\boldsymbol{a}}_{k,h,\Theta} = \theta$ for all $h \in N(g)$ and all k such that $|a| - t \leq k \leq |a|$, then

$$\tau_{\mathcal{E},g}(\overline{\boldsymbol{a}})(E(\theta)) = 1$$

Finally, a testimonial norm is said to be **sensitive** if for all expert networks $\mathcal{E}$, all agents g in the network, and all $\Theta \in \Theta_{\mathcal{E}}$, there is some $\epsilon > 0$ and some $J \subseteq PN(g, \Theta)$ such that for all $\overline{\boldsymbol{a}} \in ((A_{\mathcal{E}-g})^{N(g)})^{<\mathbb{N}}$, if $\overline{\boldsymbol{a}}_{|\overline{\boldsymbol{a}}|,h,\Theta} = \theta$ for all $h \in J$, then:

$$\tau_{\mathcal{E},g}(\overline{\boldsymbol{a}})(E(\theta)) > \epsilon$$

By construction, the six testimonial norms in the body of the paper are time-homogeneous, Markov, stable, and sensitive.

A **group testimonial norm** (or GTN for short) is a proper class function from expert networks to vectors of testimonial norms for each agent in the network. A GTN is said to be **pure** if it is a constant function; it is said to be **mixed** otherwise. We would

5.7 Scientific Communities, Probabilities over Answer Sequences, and More on Convergence

A **scientific network** is a pair $S = \langle \mathcal{E}, \langle \tau(g) \rangle_{g \in G} \rangle$ consisting of an expert network $\mathcal{E}$ and an assignment of testimonial norms $\tau(g)$ to each agent g in the network. We abbreviate $\tau(g)_{\mathcal{E},g}$ by $\tau_{\mathcal{E},g}$ below, as no confusion will arise.

Define $W_{\mathcal{E}} = \{W_g : g \in G\}$, and let $\overline{w} \in \times_{W \in W_{\mathcal{E}}} W$ be the true state of the world for all questions faced by agents in the network. Recall, in a given world, an agent's methods induces a probability measure of answer sequences *within* her area of expertise. Moreover, testimonial norms specify the probability that agents will assign to various answers *outside* their area of expertise. Therefore, given a scientific community S and world $\overline{w} \in$

$\times_{W \in W_{\mathcal{E}}} W$, one can define a probability measure $p_{\overline{w},S}$ over infinite sequences of answers for the entire network $((A_{\mathcal{E}})^G)^{\mathbb{N}}$ (where, the events are those in the product algebra). Defining the measure $p_{\overline{w},S}$ is rather involved, but it is tedious. So the details are omitted.

The measure $p_{\overline{w},S}$ allows us to characterize the asymptotic reliability of different GTNs in much the same way we characterized the reliability of methods. To do, we first introduce two more definitions. First, an expert network $\mathcal{E}$ is said to be **expert connected** if, for every question $\Theta \in \Theta_{\mathcal{E}}$, every connected component of G contains some agent g such that $\Theta_g = \Theta$. See **Figure 5** for an example of an expert-disconnected network; the right network is expert-disconnected because one connected component does not contain a red expert.

Second, say an expert network $\mathcal{E}$ is **s.a.s methodologically convergent** (or consistent, uniformly consistent, etc.) if the methods employed by each agent are s.a.s (respectively, or consistent, uniformly consistent, etc.). Given $\overline{w} \in \times_{W \in W_{\mathcal{E}}} W$, let $a(\overline{w}) \in A_{\mathcal{E}}$ be the unique answer sequence such that $\overline{w} \in a(\overline{w})$. That is, $a(\overline{w})$ is the sequence of true answers to every question if $\overline{w}$ describes the true state of the world. Say a GTN to be **s.a.s testimonially convergent** (for short, s.a.s. t-convergent) if for all scientific networks $S = \langle \mathcal{E}, \langle \tau(g) \rangle_{g \in G} \rangle$:

$$p_{\overline{w},S}((\forall g \in G)\overline{\boldsymbol{a}}_{n,g} = a \text{ for large } n) = 1$$

whenever $\mathcal{E}$ is an expert connected, s.a.s. methodologically convergent network. The definitions of convergence, uniform consistency, and so on are similar.

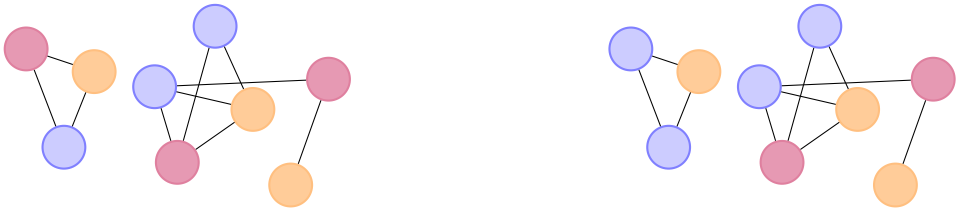


Figure 5: An expert-connected network vs. a expert-disconnected network

Notice that, in the convergence definitions for testimonial norms, we require the expert networks to be expert-connected and convergent in some way. We make these requirements because our goal is to characterize the reliability with which testimonial norms *transfer* information and not the reliability of the methods used to reach conclusions from data in the first place. Expert-connectedness ensures that each agent has (at a minimum) indirect access to data concerning every question under investigation.

5.8 Proofs of Theorems

Because the measure $p_{\overline{w},S}$ was not explicitly constructed, the proofs of the theorems below can only be sketched. Contact the author if greater detail is desired.

Given an expert network $\mathcal{E}$, define $\boldsymbol{a}(\overline{w}) \in (A_{\mathcal{E}})^G$ to be the vector representing the state in which all agents believe $a(\overline{w})$ (i.e., they believe the total truth). In contrast, let $\boldsymbol{A}(\overline{w}) = \{\boldsymbol{a} \in A_{\mathcal{E}}^G : w_g \in \boldsymbol{a}_{g,\Theta_g}\}$ to be the set of belief vectors for the network in which every agent holds a true belief *in her own area of expertise.* Next, define by recursion:

$$\begin{aligned} E_0(\overline{w}) &= \{\overline{\boldsymbol{a}} \in (A_{\mathcal{E}}^G)^{\mathbb{N}} : (\forall n \in \mathbb{N})\overline{\boldsymbol{a}}_n \in \boldsymbol{A}(\overline{w})\} \\ E_{n+1}(\overline{w}) &= \{\overline{\boldsymbol{a}} \in (A_{\mathcal{E}}^G)^{\mathbb{N}} : (\forall k \geq n+1)\overline{\boldsymbol{a}}_k \in \boldsymbol{A}(\overline{w})\} \setminus E_n(\overline{w}) \end{aligned}$$

So E_n is the event that n is the first stage at which every agent has converged to the correct answer in her own area of expertise. Finally, let $X_n : (A_{\mathcal{E}}^G)^{\mathbb{N}} \to A_{\mathcal{E}}^G$ be the function $\overline{\boldsymbol{a}} \mapsto \overline{\boldsymbol{a}}_n$ that represents the beliefs of *all* agents, to *all* questions (including those in their area of expertise) on stage n.

Lemma 2 *Let $S = \langle \mathcal{E}, \langle \tau(g) \rangle_{g \in G} \rangle$ be a scientific community and $\overline{w} \in \times_{W \in W_{\mathcal{E}}} W$. Suppose that $\mathcal{E}$ is methodologically s.a.s. convergent and that $\tau(g)$ is time-homogenous and Markov for all $g \in G$. Then $\langle X_n, E_n(\overline{w}) \rangle_{n \in \mathbb{N}}$ is a pc-Markov process over state space $\boldsymbol{A}(\overline{w})$ with respect to $p_{\overline{w},S}$.*

Proof Sketch: Since $\mathcal{E}$ is s.a.s. convergent, it follows that $p_{\overline{w},S}(\cup_{n \in \mathbb{N}} E_n(\overline{w})) = 1$. Notice the $E_n(\overline{w})$'s are disjoint by definition. Next, notice that conditional on E_n, each agent's beliefs change only *outside* her area of expertise at every stage $k \geq n+1$ (as, by definition of E_n, agents' have converged to the true answer within their area of expertise at stage n). Hence, agents' beliefs at any stage $k \geq n$ depend only upon the randomness of testimonial norms, and not upon the randomness of data. Since the testimonial norms are Markov and time-homogeneous, the vectors of all agents beliefs at stages past n, represented by $\langle X_k \rangle_{k \geq n}$, form a time-homogeneous, Markov process conditional on E_n as desired.

□

Theorem 13 *Let $S = \langle \mathcal{E}, \langle \tau(g) \rangle_{g \in G} \rangle$ be a scientific community and $\overline{w} \in \times_{W \in W_{\mathcal{E}}} W$. Suppose that $\mathcal{E}$ is methodologically s.a.s. convergent and that $\tau(g)$ is time-homogenous, Markov, stable, and sensitive for all $g \in G$. Then $\langle X_n, E_n(\overline{w}) \rangle_{n \in \mathbb{N}}$ is a uniformly absorbing pc-Markov process with respect to $p_{\overline{w},S}$, where the unique absorbing state is $\boldsymbol{a}(\overline{w})$.*

Proof Sketch: By the previous lemma, $\langle X_n, E_n(\overline{w}) \rangle_{n \in \mathbb{N}}$ is a pc-Markov process. For all agents g in the network, one can use sensitivity to show, by induction on g's length n from a Θ-expert, that there is some non-zero probability that g will believe an answer θ to Θ exactly n many stages after all the most proximate Θ experts to g believe θ. Again, using stability and induction, one can show that, for all natural numbers n and k, if g believes θ on stage n and all the most proximate Θ experts to g continue to believe θ for k stages, then g will believe θ on stage $n + k$. Since the network is s.a.s. convergent, this suffices to show that true beliefs will eventually propagate through the entire network. By the stability and Markov property of the testimonial norms, the network will be absorbed in this state.

□

5.9 Modeling Miscommunication

Suppose for the remainder of this section that the each agent's question has only two candidate answers like the running example in the text. In order to model miscommunication, one needs only to alter the definition of the measure $p_{\overline{w},S}$, so that, on each stage of inquiry, for all her neighbors, an agent reports the answer other than the one she believes with some fixed probability $\epsilon > 0$. Call the measure induced by this process $p_{\overline{w},S,\epsilon}$.

Theorem 14 *Let $S = \langle \mathcal{E}, \langle \tau(g) \rangle_{g \in G} \rangle$ be a scientific community and $\overline{w} \in \times_{W \in W_{\mathcal{E}}} W$. Suppose that $\mathcal{E}$ is methodologically s.a.s. convergent and that $\tau(g)$ is one of the six testimonial norms considered in the body of the paper for all $g \in G$. Then $\langle X_n, E_n(\overline{w}) \rangle_{n \in \mathbb{N}}$ is a uniformly regular pc-Markov process (over state space $\boldsymbol{A}(\overline{w})$) with respect to $p_{\overline{w},S,\epsilon}$.*

Proof Sketch: By the same reasoning as above, the process is a pc-Markov process. So it suffices to show it is regular. In fact, since each of the six testimonial norms has a memory of length one, the process can transition from any state in $\boldsymbol{A}(\overline{w})$ to another in exactly one step. To show this, we show that any agent's belief, with respect to any question, changes with positive probability and stays the same with positive probability. This suffices because there are only two answers to a question.

Consider a fixed agent g and a fixed question Q. Since there are two possible answers, the agent's belief with respect to Q can be represented by a 0 or 1, and her neighbors beliefs with respect to Q can be represented by a binary vector $\boldsymbol{a}$. Now each of the six testimonial norms has the following property: there are binary vectors $\boldsymbol{b}_{stay}$ and $\boldsymbol{b}_{change}$ such that, (i) if g thinks her neighbors' beliefs with respect to Q are represented by $\boldsymbol{b}_{stay}$, then g's beliefs with respect to Q will remain the same with positive probability, and (ii) if g believes her neighbors' beliefs are represented by $\boldsymbol{b}_{change}$, then g's beliefs will change with positive probability. For instance, if g is a Reidian who currently believes 0, then the constant vector containing only zeros is one example that could be $\boldsymbol{b}_{stay}$, and the constant vector containing only ones is one example of $\boldsymbol{b}_{change}$.

Let n be the number of entries in the vector $\boldsymbol{a}$ (which g's neighbors' current beliefs) differs from the vector $\boldsymbol{b}_{stay}$. Then, by the definition of miscommunication, the probability that g will think her neighbors believe $\boldsymbol{b}_{stay}$ is ϵ^n. By definition of $\boldsymbol{b}_{stay}$, if g believes her agents believe $\boldsymbol{b}_{stay}$, then g will retain her belief with some positive probability δ. So the probability that g's belief will stay the same is at least $\delta \cdot \epsilon^n$, which is positive. A similar argument shows that g's belief changes with respect to question S with positive probability.

□

Theorem 15 *Let $S = \langle \mathcal{E}, \langle \tau(g) \rangle_{g \in G} \rangle$ be a scientific community and $\overline{w} \in \times_{W \in W_{\mathcal{E}}} W$. Suppose that $\mathcal{E}$ is methodologically s.a.s. convergent and that $\tau(g)$ is one of the six testimonial norms considered in the body of the paper for all $g \in G$. Then the error rate of the network approaches a fixed value.*

Proof Sketch: By the previous lemma and Theorem 12, there is a fixed distribution that the Markov chain will be in any given state (i.e. a specification of beliefs for every agent in the network) in the limit, regardless of the initial beliefs of each agent. In each such state, there is some number of erroneous beliefs. The error rate is the expectation of error relative to this limiting distribution.

□

Appendix B - Data Summary and Statistical Tests

B1. Convergence Times for Pure GTNs

The following tables summarize the results of a one-way, random effects analysis of variance of the convergence times of the four pure GTNs on networks of 50 and 100 agents; similar results were obtained for networks of 60, 70, and 80 agents and the (analyzed) data can be obtained from the author. Here, imagine each *network* is subjected to a "treatment" of a pure GTN. I list the mean and variance of the convergence times of each of the four GTNs when the number of agents and difficulty of the problem is held constant.

When the number of agents and difficulty of the problem are held fixed, let n be the number of networks subjected to each of the pure GTNs. For brevity, it will be helpful to number the GTNs; let the numbers zero through three represent Reidians, e-trusters, proximitists, and majoritarian proximitists respectively. So μ_0, μ_1, μ_2, and μ_3 represent the mean convergence times of Reidians, e-trusters, proximitists, and majoritarian proximitists respectively. Let μ be the average of the μ_i's. Finally, let X_{ij} be the j^{th} sample point in the i^{th} GTN. Then the mean sum of treatment residuals (MSTR) and mean squared error (MSE) are defined as follows:

$$\begin{aligned} MSTR &= \frac{\sum_{i\leq 4}(\mu_i - \mu)^2}{3\cdot n} \\ MSE &= \frac{\sum_{j\leq n}\sum_{i\leq 4}(X_{ij} - \mu_i)^2}{4\cdot(n-1)} \end{aligned}$$

Then, under the null hypothesis that $\mu_i = \mu_j$ for all i, j, the statistic $\frac{MSTR}{MSE}$ has an F distribution with $3, 4\cdot(n-1)$ degrees of freedom.

50 Agents, Difficulty 0:

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	109	69.43	51.48	50.67	182
Variance	6620.03	1333.49	694.73	678.12	

MSTR	MSE	F-Statistic	P-value	Decision
134295.9939	2347.504162	51.48	$3.4 \cdot 10^{-33}$	Reject null

50 Agents, Difficulty 1:

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	2188.35	2161.43	2143.99	2075.63	182
Variance	3494459.80	3445192.34	3432294.65	2523214.5	

MSTR	MSE	F-Statistic	P-value	Decision
421153.10	3263114.68	0.13	.94	Do not reject null

50 Agents, Difficulty 2:

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	11717.40	11667.93	11669.60	12061.74	182
Variance	286146690.9	286315861.7	286325510.6	181769916.6	

MSTR	MSE	F-Statistic	P-value	Decision
6554253.48	261576729.7	0.03	.99	Do not reject null

100 Agents, Difficulty 0:

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	58.84	40.2	38.5	38.18	547
Variance	441.15	406.13	422.44	397.57	

MSTR	MSE	F-Statistic	P-value	Decision
54477.94	416.58	130.77	$6.04 \cdot 10^{-6}$	Reject null

100 Agents, Difficulty 1:

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	1631.04	1619.65	1619.32	1626.88	547
Variance	1525950.25	1523186.23	1527091.808	1507969.15	

MSTR	MSE	F-Statistic	P-value	Decision
17961.92	1523835.17	.01	.99	Do not reject null

100 Agents, Difficulty 2:

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	9500.54	9538.55	9519.88	9773.41	547
Variance	118741818.6	130970540.1	130938650.4	103751751.5	

MSTR	MSE	F-Statistic	P-value	Decision
8937190.30	121322486.3	0.07	0.97	Do not reject null

B2. Error Rates for Pure GTNs in Presence of Miscommunication

This appendix analyzes the effect of pure GTNs on the asymptotic error rate. The tables below list the **total number** of false beliefs in the network (on average) asymptotically; to obtain the error rates (i.e. the **frequency** of false beliefs), one should divide the numbers in the "mean" row by five times the number of agents (as each agent had beliefs about five questions).

The null hypothesis is that the four pure GTNs possess identical average error rates when the number of agents and probability of miscommunication is held fixed. The methodology is exactly the same as the previous two appendices. Again, only some of the results are reported because the null hypothesis was rejected every time, and the ordering of magnitudes of the error rates was identical for all networks of a fixed size and all problems of similar difficulty.

50 Agents, 1% Miscommunication:

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	33.40	16.61	4.57	3.46	434
Variance	231.63	153.09	2.26	2.6	

MSTR	MSE	F-Statistic	P-value	Decision
84244.02	130.16	647.24	0	Reject null

50 Agents, 40% **Miscommunication:**

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	98.20	93.32	92.19	85.02	434
Variance	1.55	10.61	9.24	36.94	

MSTR	MSE	F-Statistic	P-value	Decision
2846.18	19.49	658.98	$2.68 \cdot 10^{-260}$	Reject null

100 Agents, 1% **Miscommunication:**

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	44.55	9.47	5.75	2.80	749
Variance	312.52	33.53	0.88	0.97	

MSTR	MSE	F-Statistic	P-value	Decision
283738.76	116.13	2443.38	0	Reject null

100 Agents, 40% **Miscommunication:**

	Radical	Moderate	Proximitist	Majoritarian	Sample size for each GTN
Mean	195.03	175.36	173.75	153.59	749
Variance	4.13	42.09	31.81	197.89	

MSTR	MSE	F-Statistic	P-value	Decision
214750.12	92.1	2331.75	0	Reject null