

Respecting the Autonomy to Reflect:
Entitlement, Trust, and Computer Testimony

Joel Jordon

A Thesis
in the
School of Graduate Studies

Presented in Partial Fulfillment of the Requirements for the
Degree of Master of Arts (Individualized Program in Humanities) at
Concordia University
Montreal, Quebec, Canada

March 2020

© Joel Jordon, 2020

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Joel Jordon

Entitled: Respecting the Autonomy to Reflect: Entitlement, Trust, and Computer
Testimony

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Individualized Program in Humanities)

complies with the regulations of the University and meets the accepted standards with respect to
originality and quality.

Signed by the final Examining Committee:

_____ Chair
Dr. Rachel Berger

_____ Examiner
Dr. Rilla Khaled

_____ Examiner
Dr. David Waddington

_____ Supervisor
Dr. Ulf Hlobil

Approved by _____
Dr. Rachel Berger, Graduate Program Director

March 11, 2020

Dean of Faculty

ABSTRACT

Respecting the Autonomy to Reflect: Entitlement, Trust, and Computer Testimony

Joel Jordon

The beliefs that we receive through computer use and our trust in their sources might be justified or not, and I will argue that they are not prima facie justified if they threaten our autonomy to reflect afterward. Drawing on Brandom, I will talk about justification in terms of being “entitled” to a belief as a social status instituted by a practice of mutual recognition. I will argue that a lack of mutual recognition may create contexts where we either do not trust, preventing beliefs from being shared, or misplace trust, threatening that social norms may be one-sidedly enforced on our thought and behavior. When mutual recognition is achieved, on the other hand, we have a prima facie entitlement to accept beliefs and trust sources because it ensures respect for our autonomy to reflect, up to having second thoughts about our beliefs and retracting our trust. In light of my analysis of entitlement as a social status, failures of entitlement in our contemporary belief sharing practices that use intermediaries like computers will be able to be identified as historically situated failures of mutual recognition among communities. I will suggest that computers and software have to be designed for, and backed by, communities in which mutual recognition can already succeed. Then they could be tools facilitating contexts in which trust between users and programmers could develop at the same time as their autonomy to form and reform their own beliefs to think and act for themselves is respected.

Acknowledgments

Thank you to Ulf Hlobil for his thoughtful guidance and insight throughout the course of this project. Thank you to Rilla Khaled and David Waddington for their encouragement and constructive feedback.

Thank you to Christine Sedge, Stephanie Cheng, Mädchen Specht, Katie Sherman, Cameron McIntyre, Shawn Huberdeau, Jax Mill, Malcolm Geary, Ben Montague, Bayonne Said, Gigi Marko, Raymond Cacciatori, Kristen Lewis, and Emma Sigsworth for keeping my mind off work with all the experiences that nevertheless continually brought me back to it and inspired it.

Extra thank you to Shawn Huberdeau for his feedback and for always weathering the melancholy and celebrating the accomplishments with me.

Thank you above all to Hang Wu for her feedback and her endless support, patience, and humor. Although she disagrees with much of it, I couldn't have done it without her.

Contents

1. The Sociality and the Autonomy of Reasoning	1
1.1. Introduction: Computers and Recognition	1
1.2. Outline of Argument	2
1.3. Methodology: A Pragmatist Perspective on Entitlement	4
1.4. The Power to Have Second Thoughts	6
1.5. Conclusion: The Autonomy to Reflect.....	9
2. Entitlement to Testimony through Mutual Recognition	10
2.1. Introduction: Explanations and Autonomy	10
2.2. Mutual Recognition.....	12
2.3. Mutual Recognition and Technology.....	14
2.4. Entitlement to Immediate Success Testimony	16
2.5. Entitlement to Immediate Failure Testimony.....	20
2.6. Conclusion: Securing Explanations	23
3. Trust and Computer Testimony	25
3.1. Introduction: Trust and Social Norms.....	25
3.2. Trust and the Autonomy to Reflect.....	27
3.3. Trust and Software Testimony	28
3.4. Trust and User Testimony	33
3.5. Conclusion: Logical Tools for Representing Explanations.....	34
4. Conclusion: The Historicity of Recognition	36
References.....	40

1. The Sociality and the Autonomy of Reasoning

1.1. Introduction: Computers and Recognition

Criminal justice decisions increasingly rely on software to predict recidivism. With software like COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), made by Equivant (formerly Northpointe), defendants respond to a questionnaire when they are booked in jail. Their answers are used to generate “risk assessment scores” that have been taken as testimony for making decisions such as assigning bond amounts and even sentence lengths. Since recidivism is reduced to a number, and an explanation of how the algorithm works to produce this number is not accessible, there is a risk that at least some of these decisions have been made in a way where users of the software simply defer to its testimony without treating it as open to further explanation. In that case, the designers of the software might have significant control over what happens to defendants. This may be a worry not just because Equivant is a private corporation whose primary interest is making a profit by selling its software, but also because the software has been shown to be biased against black defendants.¹

Deferral to testimony while not treating it as open to further explanation is not a phenomenon unique to contexts in which intermediaries are used. A failure to treat testimony as open to further explanation is already likely to happen in a criminal justice context as long as bias and other factors keep some people from being heard. The defendant is often unaware at the time that their answers to the questionnaire will be input into COMPAS, and may only become aware of their risk assessment score in the course of their trial, if at all. COMPAS just serves to facilitate ignoring the direct explanations of the defendant and other sources representing them. We will ultimately have to consider its design and use as a consequence of historically situated social relations. Specifically, I will argue that we are not *prima facie* justified in the testimony we get from COMPAS if we don't treat it as open to further explanation because this means it was not backed by a social practice respecting our autonomy to reflect.² This social practice is mutual

¹ I take the details about how this software is used and the results of its assessments from Castro 2019, Angwin et al. 2016, and Equivant's own “Practitioner's Guide to COMPAS Core” (<http://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf>).

² For a belief to be *prima facie* justified means for it to be defeasibly justified. There are many ways a belief may become more justified or unjustified afterward. I will further discuss this later. But the reason that *prima facie* justification in testimony is important on my account in spite of just being *prima facie* has already been identified: if some testimony is *prima facie* justified, our autonomy to reflect was not threatened.

recognition. I will argue that a lack of mutual recognition may create contexts where we either do not trust, preventing beliefs from being shared, or misplace trust, threatening that social norms may be one-sidedly enforced on our thought and behavior.

The answer to the problems I raise will not be to avoid computers altogether—often, our autonomy depends on listening to others, responding when we are asked to explain ourselves, and engaging in the various other normative social activities that help us to integrate new beliefs and reflect on the beliefs that we have, activities that are in many cases today unavoidably mediated by computers. But I will show that computers and software have to be designed for, and backed by, communities in which mutual recognition can already succeed. Then, I will argue, they could be tools facilitating contexts in which trust between users and programmers could develop at the same time as their autonomy to form and reform their own beliefs to think and act for themselves is respected. This argument will show how mutual recognition is central to testimony in general, as a social practice whose success or failure is grounded in particular historical contexts.

1.2. Outline of Argument

In this paper, I will assume that “computer testimony” refers to any belief we receive through a computer, whether it comes from another user or, simply because computers and software are human artifacts, from a programmer, engineer, or institution.³ These beliefs might be justified or not, and I will talk about justification in terms of being “entitled” to a belief, drawing on Brandom. Section 1.3 describes entitlement as a social status instituted by a social practice of mutual recognition, while section 1.4 gives an account of our power to have second thoughts to show what it may mean for our autonomy to reflect to be respected when there is mutual recognition with the source behind some testimony. These sections are preliminaries necessary for the accounts of entitlement to testimony and trust given in the two main parts of the paper (sections 2 and 3).

³ I take “testimony” to include both beliefs in assertive propositional form and content intended to be formulated in this way. I think the latter significantly broadens the applicability of the account that follows because it can include responses to our queries, as well as information from diagrams, both of which are often involved in computer use. We just have to be sure to clearly demarcate such testimony from beliefs only reached by private reasoning or empirical judgment. Although it isn’t always easy to draw this distinction, I take it that there always is one. I will return to this distinction in section 2.4 when discussing how we might seem to fulfill the conditions I give for achieving mutual recognition with the source of some testimony through an intermediary, but where mutual recognition seems improbable. The beliefs at issue, I will argue, are not testimony.

In the first main part, I will argue that prima facie entitlement to testimony follows from respect for our autonomy to reflect. This respect is guaranteed by mutual recognition. *Mutual recognition*, in a testimonial context, means recognizing a rational source behind the testimony and being recognized as rational in turn. I will draw a distinction between testimony for which there is mutual recognition immediately with the source (what I will call *immediate success testimony*) and testimony for which there is an immediate failure of mutual recognition with the source (*immediate failure testimony*). Such a twofold distinction is helpful for explaining the variety of ways in which we get beliefs through computer use.⁴ We can be prima facie entitled to immediate success testimony just by treating it as open to further explanation, but we have to make up for immediate failures of mutual recognition by actually receiving an explanation for the testimony and treating it as an explanation for it. An *explanation* will refer to a reason or reasons that can give further understanding of a belief, specifically by allowing us to use it as a conclusion in some inference or inferences. The *understanding* that we have of a belief in general will refer to what we can do with it, in the sense of all of the inferences that we can make with it. A failure of mutual recognition with some testimony means that it is not open to further explanation or just not treated by us as open to further explanation, which threatens our autonomy to reflect even if we have some understanding of it.

After giving my account of prima facie entitlement, the second main part of my paper will show how a lack of mutual recognition creates contexts in which we either do not trust, preventing beliefs from being shared, or where we misplace trust, threatening the one-sided enforcement of social norms. I will give examples of testimony that users get from contemporary computer software to show how this may happen with testimony that is not open to further explanation or just not treated by them as open to further explanation, which would threaten their autonomy to reflect if they did trust the source. Since an immediate failure of mutual recognition can be overcome if an actual explanation is provided, I will use my account to support the view, found in debates on the opacity of algorithms and AI software, that software should give

⁴ I am assuming that computers are not themselves rational by restricting my inquiries to existing computers, because I am concerned with the contemporary uses to which computers are put. It is true that I am concerned with this for future-oriented reasons, but if we want to prescribe how computers should be, I think we should start here. If it turns out that computers can reason in the future, then we might be able to recognize them as rational sources directly while they recognize us as rational in turn. I think that this paper's reflections on orienting computer use and design around a practice of mutual recognition that is central to testimony in general will still be relevant in that case.

explanations. But I will conclude my paper by arguing that we have to see failures of mutual recognition, and the design of the software through which we experience them, as social failures among different historical communities. They may occur with some recipients of testimony but not others, depending on whether they recognize and are recognized by the source. Given that much of our digital and network infrastructure has its origins in military, corporate, and other institutional sources, there is, I will show, a question of whom these sources recognize and are recognized by.

1.3. Methodology: A Pragmatist Perspective on Entitlement

In this section, I will argue that a concept of entitlement as a social status can capture how we treat ourselves and others in our practices of belief sharing. This notion of entitlement has to first be explained so that I can use it in my account of prima facie entitlement to testimony. I will introduce it here to show how it can be used to intervene in debates on entitlement to computer testimony.

Throughout this paper, I will be using Brandom's definition of "entitlement" to a belief as a social status giving us authority to it. It is vindicated with an explanation for the belief it gives us authority to. This notion of entitlement can be clarified by comparing it with one not conceived as instituted by us, such as that found in Burge 2009 (272). According to the latter, we become entitled if a belief achieves its purpose, in the same way as a competent faculty does.⁵ On that view, a belief would be seen as constitutively aimed at truth. This can explain how we become entitled to beliefs when they seem to get things right, but it is hard to further specify what this means using this concept of entitlement. If we understand entitlement as instituted by us, on the other hand, then it can be understood as a social status conferring authority, and we can understand instances where it is not achieved as failures of whatever practice instituted it. Entitlement as a social status instituted by us can capture how we treat ourselves and others as having or not having a *right* to different beliefs. And it can explain how social norms can be made and enforced through beliefs. Explaining entitlement in these terms involves taking a step

⁵ Burge discusses entitlement in contrast to justification as two kinds of warrant for a belief. A justification for him is a reason for a belief, which may be closer to what Brandom means by entitlement. I talk about his entitlement in this paper because he thinks we have a prima facie entitlement to testimony, not a prima facie justification in it. It might be replied that I am reconstructing an account of how testimony might be prima facie justified in Burge's sense. But while Burge's justification entails having reasons, it is, like his entitlement, not a social status and similarly lacks the power to explain how we treat ourselves and others in our actual practices of belief sharing.

back into what Brandom 2008 has called a *pragmatic metavocabulary* that can say what is done in, for example, becoming and being entitled to a belief. Then I can describe how in particular it succeeds or fails to be instituted in cases involving the use of computers.

This pragmatic perspective on entitlement can be used to intervene in the literature on computer testimony. Literature on entitlement to testimony in general splits along the lines of antireductionism, the view that we can have a prima facie entitlement just in virtue of testimony's status as testimony, and reductionism, the view that our entitlement to testimony reduces to other factors that establish the testimony's reliability. This debate also occurs specifically in literature on entitlement to testimony from computer-assisted mathematical proof. Burge 1998, Wong 2012, and Teller 1980 argue against Tymoczko 1979 and Kripke 1980 to defend the idea that at least some entitlement to computer testimony is a priori as opposed to empirical. This is done on the basis of Burge 1993's antireductionist "acceptance principle," which states that we have an a priori entitlement to testimony when we have seeming understanding of an assertion, because this implies that we can recognize a rational source behind the testimony. The distinction between a priori and empirical entitlement is not the essential one for my purposes, but I will hew closely to Burge's antireductionist position by arguing for principles governing prima facie entitlement similar to the acceptance principle, just with stricter requirements involving explanations. In explicating these principles, however, I will argue that entitlement does not involve merely recognizing a rational source, since this cannot explain the kind of normativity that beliefs exhibit in our practices of belief sharing. My principles will use Brandom's different concept of entitlement as a social status, which will have to be kept clearly distinct from Burge's concept.

Drawing on Brandom 2019, I will argue that entitlement as he conceives it must be instituted by a two-sided form of recognition that consists in recognizing a rational source who recognizes us as rational in turn. In a testimonial context, mutual recognition is achieved, I will argue, when testimony is treated as open to further explanation or actually comes with an explanation that we treat as an explanation for it. Entitlement is only a social status instituted by a social practice like this one. We of course hold many beliefs without expressing them, but on this account, unless we had previously instituted an entitlement for them by a practice of mutual recognition, we would not be entitled to them. One way of justifying this might be to follow Brandom 1994 (153–54) in seeing all our beliefs as essentially assertions. On this view, all our

private beliefs are possible assertions. We might say they are of the form of claims that we could become entitled to once they are expressed and met with a successful instance of mutual recognition. If we understand beliefs in this way, then it is *only* through instituting at least a prima facie entitlement to testimony that we become entitled to any of our beliefs. A source expressing their belief in a way that would prima facie entitle us to it entitles them to it too. The recipient's prima facie entitlement can be overridden if they have reason to disagree with the testimony, but the source's status of entitlement can be transferred as long as there is mutual recognition and the recipient does not have stronger reasons counting against the belief. If they do, they could be prima facie entitled to their incompatible belief if it was similarly met by a successful instance of mutual recognition. But it is only by achieving mutual recognition and instituting a prima facie entitlement that source and recipient can become entitled to any of their beliefs at all.

As will be borne out in my account, entitlement understood as a social status instituted by a practice of mutual recognition can make sense of how we treat ourselves and others in our practices of belief sharing. And it can be used to explain how social norms may be made and enforced by our beliefs. If, as I will argue in section 2, mutual recognition ensures respect for our autonomy to reflect, then it is by mutual recognition that we can guide the creation of norms for our thought and behavior. When, on the other hand, we accept a belief without it being backed by a practice of mutual recognition—without our being or becoming entitled to it—it is being given authority that it does not have, and it could act to one-sidedly enforce social norms on our thought and behavior.

1.4. The Power to Have Second Thoughts

I conclude the introduction with a description of a kind of inferential content of our beliefs that I call its incompatibility content and a discussion of how this content is used in the reflection involved in having second thoughts. This is still preliminary to talking about testimony itself, but it is important for that account because it shows what it means to have the autonomy to reflect on our beliefs. My paper will proceed on the basis of the idea that having second thoughts is one high-level, but generally available kind of such reflection. I will argue that we have the ability to have second thoughts about any beliefs at all that we have some understanding of—not

just to add or remove from our stock of beliefs, but to put incompatible beliefs into question by setting them against each other and inferring some over others.⁶

I will first clarify some concepts that I use in my account of our ability to have second thoughts. To recall, the understanding that we have of a belief in general refers to what we can do with it, in the sense of all of the inferences that we can make with it. What I will call a belief's *incompatibility content* refers to the understanding that we have of a belief that specifically allows us to make the inferences with it that are involved in having second thoughts. By incompatibility, I mean more than mere logical inconsistency, but rather a kind of incompatibility describable in terms of rules of material inference, which is why I think it can be referred to as part of the content of the beliefs. I think that this incompatibility content is guaranteed if having some understanding of beliefs always involved having some beliefs in the denial of beliefs that they are incompatible with. These might be implicit beliefs, as long as they are discoverable. By *implicit beliefs*, I mean claims that we in practice affirmed, but did not assert (either privately or publicly), while taking up the beliefs that we hold explicitly. They are beliefs that we can potentially make inferences with if we first discover them through deductive reasoning. What I have in mind is the following, based on Peirce 1880 and 1885's account of logic as logical criticism that exposes what he calls "leading principles":

All deductive reasoning, even simple syllogism, involves an element of observation; namely, deduction consists in constructing an icon or diagram the relations of whose parts shall represent a complete analogy with those of the parts of the object of reasoning, of experimenting upon this image in the imagination, and of observing the result so as to discover unnoticed and hidden relations among the parts. (Peirce 1885, 182)

The example I will use to elucidate this is based on Peirce 1880 (18–19). Say we have the following argument:

All computers are machines.

All machines are irrational beings.

∴ All computers are irrational beings.

Peirce would argue that the leading principle of this is *nota notae est nota rei ipsius* ("the predicate of the predicate is the predicate of the subject," or more simply, predicates are

⁶ I draw on Macbeth 2014, 184–88, who herself draws on Kant, for the concept of this robust ability to have second thoughts. That we can put any of our beliefs into question, just not all at once, also has precedent in Sellars 1956.

transitive). Now say that we sufficiently understand the belief “All computers are machines” to make the inference involved in this argument. This means that it is at least inferentially articulated by the two other explicit beliefs in the argument. But my idea is that what Peirce identifies as this argument’s leading principle might also be an implicit belief we laid claim to in making this inference. Someone else might have a different understanding of these beliefs that allows them to make the same inference without this implicit belief. By hypothesis, all of the explicit beliefs involved in their understanding that enables them to make this inference are the same for this person. But they might be able to make this inference with a rule other than one prescribing general validity to its form. It often isn’t easy to “observe” our own reasoning after the fact, let alone someone else’s, to make discoveries of the implicit beliefs involved in it. Nevertheless, I define an implicit belief as any claim implicit in the reasoning involved in reaching the conclusion that we are explicitly committed to.

I have introduced this idea of implicit beliefs to use them in my account of why we can have second thoughts about any beliefs we have some understanding of at all. We have at least a degree of understanding of our beliefs when we can make some inferences with them; this might mean that they are inferentially articulated by some explicit beliefs, but they have to be articulated at least by some discoverable implicit beliefs. To say that we can have second thoughts about any belief we have some understanding of means that a minimally understood belief at least has incompatibility content. I will try to justify this assumption by defining entailment in terms of incompatibility, following Brandom 2019: “ Q is a consequence of P just in case everything materially incompatible with Q is materially incompatible with P ” (144).⁷ For him, the reason for pursuing the order of explanation that puts incompatibility before consequence relations is that it gives a plausible account of how modality is built into our inferential practices from the start.⁸ It would explain how being able to treat any of our beliefs

⁷ The same definition is given in Brandom 2008, 121.

⁸ The other order of explanation, which Brandom identifies in the Tarskian tradition of extensional logic, can still be used to justify our ability to have second thoughts about any belief that we have some understanding of. This may be more appealing to those who want to take consequence relations rather than incompatibility as primary. Assume that our understanding of a belief does not include an explicit belief in the denial of an incompatible belief for some belief, but only a reason for holding it or a consequence that follows from it. Just being able to make inferences with these beliefs does not, on this order of explanation, mean that we have some understanding of what the belief is incompatible with. But we *could* reject the reason or the consequence. Doing this, however, we would precisely be making explicit some belief incompatible with it, namely, the belief in the denial of the reason or consequence. Then we should be able to have second thoughts about the belief in light of that belief. Notably, this would mean that for beliefs for which our understanding does not already include incompatibility content, the incompatible beliefs we

that we have some understanding of as possible or necessary is foundational and discoverable if not already explicit. For example, say we have the understanding of the belief “Facebook must be valued by its users” sufficient to use it as the conclusion to the following inference:

If a service meets a demand in the free market, it must be valued by its users.

Facebook meets a demand in the free market.

∴ Facebook must be valued by its users.

Then we must also have at least implicitly laid a claim to, and be able to discover, the denial of something that we take to be impossible to hold if these beliefs hold, in the form of some belief or beliefs that they are incompatible with. Say that one of these beliefs is “Facebook manipulates its users.” It seems plausible that we would have at least implicitly claimed, and be able to discover, the denial of a belief like this one (but maybe not this one in particular) if having some understanding of what beliefs are incompatible with is prior to being able to use them in inferences. And if we can discover implicit incompatible beliefs for any of our beliefs that we have some understanding of, then we can have second thoughts about any of our beliefs that we have some understanding of. It guarantees that we can find some incompatibility content for our beliefs for which we only have an explicit understanding of their relation to other beliefs as a consequence or reason.

Reflecting on this incompatibility content is all that having second thoughts consists in. Inferring one belief over another based on their incompatibility affirms that it is necessary that only it may hold, or that it is not possible for both to hold. For example, we might make the following inference:

Facebook must be valued by its users.

Facebook manipulates its users.

∴ Facebook manipulates its users (denying that Facebook must be valued by its users).

1.5. Conclusion: The Autonomy to Reflect

If we have a different understanding of our beliefs, we might have second thoughts about them in light of different beliefs, and the result of having second thoughts might be different.

can use to have second thoughts about it are restricted to the formal negation of our reasons for holding it or what follows from holding it. But this just reflects the centrality of formal rules of inference for this tradition of logic, in contrast with Brandom’s inferentialism and the traditions that inspire it with their emphasis on material rules and the defeasibility of reasoning.

Inferring “Facebook manipulates its users” depends on it being found in the incompatibility content of what we are having second thoughts about. Having second thoughts, then, is something that can be done differently according to the nature and extent of our understanding of each of our beliefs. This is a key point for the discussion to come because I argue that we are *prima facie* entitled to testimony when our autonomy to reflect in general is respected. This entails respect for our autonomy to reflect not just for the testimony, but also for all the beliefs involved in our understanding of it. If we misplace trust in a source with whom our autonomy to reflect is threatened, then social norms might be one-sidedly enforced not just by the testimony, but also by the other beliefs involved in our understanding of it. If, for instance, in the example I gave our autonomy was threatened so that we *couldn't* have second thoughts about “Facebook must be valued by its users,” then we might also be unable to have second thoughts about one of the beliefs that I originally identified as belonging to our understanding of that belief, “If a service meets a demand in the free market, it must be valued by its users.” My account remains agnostic on whether all or just some of our beliefs are (or might potentially be) norms for our thought and behavior and what exactly counts as a “social” norm. But if our autonomy to reflect is threatened, then there is at least a risk that social norms may be one-sidedly enforced, whether explicitly by the testimony or implicitly by the other beliefs involved in our understanding of it. If, on the other hand, we retain the ability to have second thoughts, then we are also able to retract our trust from a source, as I will show in section 3. This protects us from the one-sided enforcement of social norms.

Although I hope to have shown that we can have second thoughts about any beliefs at all that we have some understanding of, my argument in what follows will be that our autonomy to reflect in this way and others may nevertheless be threatened when we do not treat testimony as open to further explanation. This is because this can make us treat ourselves just as if we didn't have understanding, likely about the testimony itself or the beliefs involved in our understanding of it.

2. Entitlement to Testimony through Mutual Recognition

2.1. Introduction: Explanations and Autonomy

In section 2, I will provide my account of *prima facie* entitlement to testimony. I will argue that we are *prima facie* entitled to testimony as long as our autonomy to reflect is

respected. In the introduction, I will discuss the role that explanations can play in securing this autonomy, up to being able to have second thoughts, through their involvement in a practice of mutual recognition.

To recall, my concept of *explanation* refers to a reason or reasons that can give further understanding of a belief specifically by allowing us to use it as a conclusion in some inference or inferences. Hills 2015's description of the abilities involved in understanding why p is closely related to the kind of understanding that we can get from explanations:

- (i) follow some explanation of why p given by someone else.
- (ii) explain why p in your own words.
- (iii) draw the conclusion that p (or that probably p) from the information that q .
- (iv) draw the conclusion that p' (or that probably p') from the information that q' (where p' and q' are similar to but not identical to p and q).
- (v) given the information that p , give the right explanation, q .
- (vi) given the information that p' , give the right explanation, q' . (663)

She calls this set of abilities having “cognitive control.” While I think a reason or reasons that can function as an explanation for a belief can grant cognitive control or something like it over that belief, what I mean by having understanding of a belief in general is less than what is involved in having cognitive control. We might have some understanding of p just by understanding it as a consequence of or reason for q or as incompatible with q . I have argued that we can have second thoughts about any belief that we have some understanding of. But explanations nevertheless play a central role in my account of prima facie entitlement to testimony because, I will argue, not treating testimony as open to further explanation may threaten our autonomy to reflect, up to having second thoughts about our beliefs. It can make us treat ourselves just as if we didn't have understanding about some beliefs, likely the testimony itself or the beliefs involved in our understanding of it.

I take our autonomy to reflect to be tied to a more basic practice of mutual recognition. We experience mutual recognition with a source of testimony, I will argue, when the testimony is treated as open to further explanation or actually comes with an explanation that we treat as an explanation for it. Under circumstances where mutual recognition fails, on the other hand, we might defer to testimony that either is not open to further explanation or, more likely, just not treated by us as open to further explanation. If we defer to testimony that is not open to further

explanation or just not treated by us as open to further explanation, then it threatens our autonomy to reflect. We might be able to protect ourselves from this threat by not accepting the belief unless we actually received an explanation for it. But as long as our autonomy *would* be threatened if we took the testimony, I take it that we are not *prima facie* entitled to it. This basically amounts to taking the position that practices of belief sharing should in general be backed by a more basic social practice of mutual recognition that ensures respect for our autonomy to reflect.

2.2. *Mutual Recognition*

In this section, I will give an overview of what mutual recognition consists in. This overview applies to mutual recognition for both kinds of testimony that I will discuss. My account of how treating testimony as open to further explanation or actually receiving an explanation that we treat as an explanation for it is sufficient for mutual recognition is saved until after I distinguish these two kinds of testimony in the following section. Here I will describe the overall structure of mutual recognition and show how it guarantees respect for our autonomy to reflect.

There are two parts to mutual recognition as Brandom 2019 (277–90) conceives it, specific and general. Applied to a context in which our entitlement to some testimony is at issue, what he calls the *specific* part would consist in a source recognizing us as having the authority (entitlement) to make ourselves responsible to (commit to) the testimony. But for the source to have the authority to specifically recognize us as capable of doing this, we must have recognized them as having the authority to institute statuses of entitlement and commitment *in general*. And for us to recognize the source as having this general authority, they must have conferred this authority on us too by recognizing us in the same way. Specific recognition thus also entails mutual recognition at a general level, because the authority to specifically recognize someone as instituting a status (entitlement in this case) with their attitude (a believing) depends on our and their recognition of each other as having the authority to institute statuses of entitlement and commitment in general.

I consider our recognition of one another as rational to consist in what happens at the general level of mutual recognition. What I call respect for our autonomy to reflect follows from our recognition of one another as rational. Our autonomy to reflect is threatened when there is

not recognition of one another as rational at the general level because it threatens that we may treat ourselves as incapable of contributing to the institution of entitlements and commitments to our own beliefs. Not treating our private beliefs as potential assertions that could be met by successful instances of mutual recognition could be, I argue, just the same as not having understanding of those beliefs in the sense that it could make them unusable in inferences. Inferring depends on our entitlements and commitments. This is not to say that we can't make inferences with entitlements and commitments that are not already successfully instituted by a practice of mutual recognition. Rather, our ability to use our entitlements and commitments may be threatened when we don't treat ourselves as capable of contributing to their institution. On the view that entitlements and commitments are social statuses, entitlements and commitments that we do not treat ourselves as capable of instituting may not be entitlements and commitments for us at all.

It isn't enough for our autonomy to reflect to be respected for us to either recognize the source as rational or be recognized as rational—in fact, on this account, neither can happen without the other. A source only has the authority to confer the authority to institute statuses on us if we also confer this authority on the source. So although it might seem sufficient for us to just be recognized as rational by the source for our autonomy to reflect to be respected, we also have to recognize the source as rational. For example, whether the source behind some software's testimony seems to have failed to recognize us as rational or we seem to have failed to recognize them as rational, *both* would actually happen: the recognition of the one as rational depends on the recognition of the other as rational. On this account, where recognition has to be co-instituted, there isn't a sense to a merely one-way success or failure of recognition. Some examples support a reading where either side might have *initiated* the failure of mutual recognition, while other examples support a reading where one side more explicitly initiates it. For example, it may not be certain whether a user initially failed to treat the source behind some software as rational or if the source behind the software failed to treat them as rational, or if both are somehow at fault, whereas in other cases it might be more clear-cut who initiated the failure, such as when someone is intentionally excluded as a member of a community. In what follows I will sometimes talk just about a person or institution failing to recognize or be recognized because it is an example like the latter, but it should be kept in mind that although the initial

failure might lie on one side in causal terms, the recognition then also cannot proceed in the other direction.

Mutual recognition in a belief sharing context follows, I will argue, from testimony either being treated as open to further explanation or an explanation for the testimony actually being produced. We are not, I argue, *prima facie* entitled to testimony just by having some understanding because there may nevertheless not be mutual recognition and our autonomy to reflect could thus be threatened. We are *prima facie* entitled to testimony if we not only have some understanding but also achieve mutual recognition, because it ensures that our autonomy to reflect is not threatened by treating ourselves just as if we did not have understanding.

2.3. Mutual Recognition and Technology

To give explicit principles for the conditions on *prima facie* entitlement to testimony that can apply to the variety of beliefs we get from computers will require drawing a distinction between testimony behind which we immediately recognize a rational source and are recognized in turn as rational (immediate success testimony) and testimony for which there is an immediate failure of mutual recognition (immediate failure testimony). I draw this distinction along the lines of whether or not some testimony is treated as open to further explanation. An immediate failure of mutual recognition should just be taken to mean that some testimony is not open or not treated as open to further explanation, even our own.

It is not the mere presence of technology that causes an immediate failure of mutual recognition, since in fact an immediate success or failure could occur with or without technology. The distinction is intended to leave open that some uses of books, phones, messaging systems, and computers and even some direct conversation could produce either. I intend an immediate success to be possible with testimony that we receive from some intermediaries, such as books. There can be an immediate success with such sources, even if they themselves do not contain further explanation, because the testimony could remain open to our own further explanation. In many cases this would just require that the testimony seem coherent and sincere and that it could therefore be taken to be the product of a rational source who intended the testimony for a recipient they treated as rational. However, it is possible to treat a book or some of its contents as incomprehensible and so for its testimony to lose its status as producing mutual recognition immediately. We often have an immediate success through

technology-mediated communication that creates conditions for mutual recognition much like those of direct conversation, such as phone calls and messages. But there can still be an immediate failure in any of these cases when we begin to require or even just think that we require replies and they are not forthcoming.

With computers, testimony generated by programming often produces an immediate failure. If further explanation requires, or if we just think that it requires, some information about how a piece of software or website works that is not available to us, then it may not be open or just not treated as open to further explanation. For example, in February 2020, the caucuses were held in Iowa for the selection of the Democratic Party's candidate for president of the United States. Vote tallying was riddled with issues, some technology related because a smartphone app was used to record and report results (Koebler, Cox, and Maiberg 2020). Someone who did not understand how the results of the caucuses were calculated, including technical details of how the app works, would likely experience an immediate failure because the testimony is not open or just not treated by them as open to further explanation. Someone who was, on the other hand, more directly involved in the caucuses, whether by voting in them or helping to tabulate the results, or who had more technical knowledge of the app, is more likely to treat the testimony as open to further explanation and therefore experience an immediate success. This begins to show how mutual recognition is relative to the user and cannot be seen as just related to the technology involved.

Testimony through a computer that is from another user, unlike testimony generated by programming, often produces an immediate success, but it can fail in the same ways as phone calls and messages can. The introduction of any of these technologies also makes it possible to confuse testimony from a rational source with generated testimony, making it more likely that we put into question the status of someone whom we initially took to be human, or find our own status questioned by others. Then we become susceptible to not treating the testimony as open to further explanation in the same way as we do with testimony generated by programming. It is similarly relative to the users involved and how they treat one another in their belief sharing practices. Mutual recognition is highly context sensitive. How users and programmers approach one another in giving and receiving testimony matters above all, but technology still plays a significant role in shaping those interactions. I will return to these issues later, after giving my

account of prima facie entitlement to testimony in the rest of section 2 and my account of prima facie entitlement to trust in section 3.

In this section, I have discussed different ways of treating testimony and alluded to the norms that might be operant with them without making explicit reference to what prima facie entitlement to testimony consists in. In the remainder of section 2, I will argue that we are prima facie entitled to immediate success testimony because we treat it as open to further explanation, but we are not prima facie entitled to immediate failure testimony unless we actually receive an explanation for the testimony that we treat as an explanation for it. This will explicitly secure mutual recognition after an immediate failure. By drawing more explicit principles for our prima facie entitlement to each kind of testimony, I will be able to show how mutual recognition and thus respect for our autonomy to reflect follow from prima facie entitlement.

2.4. Entitlement to Immediate Success Testimony

In this section, I will describe our entitlement to testimony for which there is an immediate success of mutual recognition. I will discuss the different requirements for prima facie entitlement to each kind of testimony in terms of principles shaped in reply to Burge 1993's acceptance principle governing our default entitlement to testimony:

Acceptance principle (AP): A person is entitled to accept as true something that is presented as true and that is intelligible to him, unless there are stronger reasons not to do so. (237)

He has elsewhere expressed this as the principle that we are entitled to accept testimony when we have seeming understanding of a presentation-as-true.

Fulfillment of this entitlement is supposed to lead to recognition of a rational source behind the testimony. I will argue, however, for a principle that I call the potential explanation principle (PEP) describing what is necessary to achieve mutual recognition, which is the only kind of recognition that would be able to institute entitlement in Brandom's sense:

Potential explanation principle (PEP): A person is entitled to accept testimony as respecting their autonomy to reflect if it fulfills the criteria of AP and they treat it as open to further explanation that is not just the explanation that they have heard it from a source.

By “fulfilling the criteria of AP,” I mean that the belief is presented as true (or presented sincerely) and is intelligible to the recipient and there are not stronger reasons not to accept the belief. The entitlement mentioned in AP itself is based on Burge’s concept of entitlement as a belief achieving its purpose and is not an entitlement that needs to be fulfilled to be entitled according to PEP. PEP does not describe what is necessary to have entitlement in this sense, but rather what is necessary for us to have a prima facie entitlement to testimony by achieving mutual recognition with the source and thus ensuring our autonomy to reflect is respected. I include that the criteria of AP have to be fulfilled because PEP isn’t meant to indicate that we can accept testimony of a belief that we have no understanding of at all as long as we expect that we could get understanding. We have to have some understanding, perhaps made up of implicit beliefs, of some sincere testimony already (the testimony has to be intelligible, as AP says). One way of thinking of it is that we must have made some claims that are involved in having any understanding of the testimony at all, but we are not entitled to those claims unless our autonomy to reflect is also respected.

PEP does not indicate where the further explanation comes from. It could come from the recipient of the testimony themselves. This would explain how we can be prima facie entitled to, for example, sincere and intelligible testimony from books that we take to be the product of a rational source who intended the testimony for a rational reader. We might expect a book to further explain some of the beliefs expressed in it, but it will always hit a limit. We can nevertheless be prima facie entitled to these beliefs if we have some further explanation for them in our own private reasoning.

According to Burge, we recognize a rational source behind the testimony just by fulfilling the criteria of AP, which means that we have some understanding of the testimony. But there is testimony that fulfills this criteria that isn’t further explainable or isn’t treated as further explainable and so, on my view, cannot actually be treated as being from a rational source. I argue that we do not actually recognize a rational source and are not prima facie entitled to this testimony because there is a stronger condition on actually recognizing a rational source that involves being recognized in turn by them. Fulfilling the conditions of PEP achieves mutual recognition and thus guarantees respect for our autonomy to reflect on the account of mutual recognition that I gave in section 2.2. If the source shared a belief in a way that we could treat it as open to further explanation, they must have *specifically* recognized us as having the authority

(being entitled) to make ourselves responsible to (commit to) it. But they only have the authority to specifically recognize us as capable of doing this because we *generally* recognize one another as rational in the sense of being capable of entitling and committing ourselves to beliefs. They must have recognized us as rational for the same reason that they specifically recognized our entitlement, that is, because they shared the belief in a way that we could treat it as open to further explanation. For our part, we must have recognized them as rational because we treated their testimony as open to further explanation. Respect for our autonomy to reflect follows from our recognition of one another as rational in this way. We are *prima facie* entitled to the testimony because we not only have some understanding of it (by fulfilling the conditions of AP) but also our autonomy to reflect is respected.

If fulfilling the conditions of PEP seems insufficient for mutual recognition in all cases, then it should be noted that the further explanation cannot be merely that one heard it from a source. Treating some testimony as open to further explanation must be an indication that we recognize one another as capable of entitling and committing ourselves to beliefs in general. But since this form of explanation just refers to the form of testimony in general rather than giving further understanding of the particular belief, it *defers* those rational abilities to a source rather than ensuring that we recognized one another as having them. We might attribute these rational abilities to the source we are attempting to recognize, but they might not have recognized us as rational if they shared the belief in a way that we could only treat it as open to the further explanation that we heard it from them. Or, in the case of secondhand testimony, we might attribute these abilities to another source altogether, and we might not have even attempted to recognize them as rational if this is the only further explanation we treat it as open to. So on this account, we are not entitled to secondhand testimony unless we treat it as open to further explanation that is not just the explanation that we heard it from a source. There has to be some minimal openness to further explanation about the belief itself rather than just some further explanation pointing to our *prima facie* entitlement to testimony in general.

In some cases, however, if this seems to be the only further explanation that we treat a belief as open to, then we may have formed an empirical judgment rather than took testimony at all. It can be difficult to distinguish testimony from nontestimony, particularly when the belief is shaped in response to an artifact made by humans. But I take it that there is always this distinction even if we can't always be sure of when to draw it. We might form a belief merely by

empirically judging it in response to an artifact, rather than taking it as testimony by treating the artifact as an intermediary through which we might recognize a rational source. For example, it is debatable whether the beliefs we form in response to Google Maps are testimony or just empirical judgments. Say that, upon using Google Maps, we could give the observational report “The bus arrives at 8:15 AM,” and the further explanation that we treat it as open to, in being able to give that report, is only “A bus schedule gave 8:15 AM as the arrival time.” I don’t think this would be sufficient for mutual recognition because the further explanation seems to just be that we heard it from a source (whoever made the bus schedule). But what appears to be an appeal to the source of some testimony may just be an appeal to the appropriateness of the conditions under which we made an empirical judgment. Our autonomy to reflect is not threatened in the latter case because there was no attempt at mutual recognition in the first place, and so there could be no failure of mutual recognition.

As Sellars 1956 has argued, empirical judgments, like our other judgments, presuppose their use in inferences, or what I have called our understanding of them. I have identified the threat to our autonomy to reflect arising from a failure of mutual recognition as a threat to treat ourselves just as if we had no understanding of some beliefs. But this is a threat that lies in our belief sharing practices, and there is no equivalent threat for forming empirical judgments. So as long we have some understanding of our empirical judgments, our autonomy to reflect on them is not threatened. Consider how having second thoughts about our empirical judgments usually involves induction. If we consistently get things wrong, we can, as long as we have some understanding of the conditions under which they were made, reevaluate those conditions. This is no different whether the conditions involve the functioning of our own faculties or a tool. If we start to get incorrect information from Google Maps, we will probably stop relying on it. Our autonomy to do this is not related to any process of recognizing rationality. The distinction between testimony and empirical judgments, although not always easy to draw, can explain some cases where we seem to be *prima facie* entitled to beliefs, but treating them as open to further explanation clearly does not seem sufficient for the entire apparatus of mutual recognition.

Not treating some testimony as open to further explanation can indicate an initial failure of recognition from either side or both. Often, the source fails to offer testimony in a way that allows us to treat it as open to further explanation, so they seem to cause the initial failure by not recognizing us. If the failure of recognition seems to initiate from our side, we will usually not be

vulnerable to threats to our autonomy because we will withhold trust from the source and not take the testimony. But it can be hard to determine in all cases which side actually initiates the failure of recognition. As I discussed in section 2.2, whoever may have caused the initial failure, recognition fails for both. And for immediate success testimony, a failure of mutual recognition is any case indicated by a failure to treat the testimony as open to further explanation.

Despite the potential for failure, PEP, like AP, is supposed to capture the ease with which we can share beliefs in a prima facie entitling way. Treating testimony as open to further explanation is a step above just having understanding, but it is not usually hard. If further explanation is actually produced, then prima facie entitlement to the explanation is also in many cases easily secured. But since these entitlements are just prima facie, we can always disagree. Describing the autonomy to reflect in terms of mutual recognition captures well the relation between prima facie entitlement and disagreement. There is at the specific level of mutual recognition an entitlement to the belief that allows us to lay claim to it, while at the general level our autonomy to reflect, up to having second thoughts about the belief, is respected. So placing mutual recognition at the bottom of belief sharing practices should not be mistaken as suggesting a global tendency toward bringing our understandings closer together. A better picture would be of how mutual recognition might establish localized communities that maintain the autonomy of their members to disagree and, perhaps, to more effectively challenge other communities who do not recognize them or with whom they disagree more deeply, as I will discuss later in the concluding section of this paper.

2.5. Entitlement to Immediate Failure Testimony

In this section, I will argue that it is still possible to be prima facie entitled to testimony for which there is an immediate failure of mutual recognition. Describing a further principle for how we can become prima facie entitled to such testimony can explain how we can make up for an immediate failure of mutual recognition that we experienced directly with another rational source or with a rational source behind testimony generated by an intermediary like a computer.

What it means for us to be unable to immediately achieve mutual recognition is that some testimony is not open to further explanation or just not treated as open to it. The clearest cases are when further explanation is actually inaccessible and so can't be treated as open to further explanation. This might include testimony of beliefs all of whose explanations are local to the

reasoning of someone else who refuses to explain themselves to us. Not understanding a belief at all is another way. But the range of immediate failure testimony is significantly broadened by the possibility that some testimony that is open to further explanation could just be treated as if it weren't. We might treat some testimony from a source that we do not recognize as rational, or who does not recognize us as rational, as not further explainable, even by us. It might happen when we begin to think that someone, when they can produce no explanation, or only an insincere or even an incoherent one, has merely heard the belief that they shared from someone else. The source of the testimony is treated like a human intermediary, rather than as rational, with respect to the belief, and we are more likely to not treat it as open to further explanation. The intervention of a technological intermediary, such as a phone or messaging system, makes it more likely that we put into question a source's status as rational, or find our own status questioned, causing us to treat their testimony as not open to further explanation. A piece of software's report of a result that it calculated to is also often not open to or not treated as open to further explanation. It might be that there is no further explanation for it because such explanations require reference to privileged information such as how the software works, or it might be that we just defer to the software's testimony without expecting that there could be further explanation for it because we do not recognize or are not recognized by the source.

We cannot or do not expect, in the contexts in which this testimony is shared, that an explanation is available. There is no *prima facie* entitlement to it under PEP. But we could still become *prima facie* entitled, I argue, if stronger conditions of actually receiving an explanation for the testimony and treating it as an explanation for it is fulfilled. If the immediate source of the testimony is themselves a rational source who has an explanation for the belief, then this would involve reestablishing mutual recognition with them, while if they are an intermediary, then we would establish mutual recognition with a rational source *through* the intermediary. I will argue for what I call an actual explanation principle (AEP) describing how mutual recognition may be explicitly secured after an immediate failure. Fulfilling its conditions *prima facie* entitles us to testimony even if the conditions of PEP are not fulfilled:

Actual explanation principle (AEP): A person is entitled to accept testimony as respecting their autonomy to reflect if it fulfills the criteria of AP, they actually receive an explanation for it, and they treat it as an explanation for it.

Mutual recognition is explicitly secured, in cases where it was not or could not be before, by fulfilling the conditions of AEP. If the source shared a belief with an actual explanation that we could treat as an explanation for it, they must have *specifically* recognized us as entitled to the testimony. And in order to do this, we again must have *generally* recognized one another as rational. I argue that, like for PEP, the source must have recognized us as rational for the same reason that they specifically recognized us, that is, in this case by sharing the belief with an actual explanation that we could treat as an explanation for it. We must have recognized the source as rational in this case because they actually offered an explanation that we treat as an explanation from them. Securing mutual recognition in this way after an immediate failure respects our autonomy to reflect. It prima facie entitles us to the testimony in the same way as it did for PEP, by ensuring not only that we understand the belief but also that our autonomy to reflect is not threatened.

AEP is carefully phrased to indicate that the recipient has to *receive* an explanation. Unlike for PEP, the explanation cannot be their own. This is because AEP secures a prima facie entitlement by making up for an immediate failure of mutual recognition through which the recipient does not treat the testimony as open to further explanation. The possibility of becoming prima facie entitled to the testimony because it was open to their own explanation was already precluded, and to have mutual recognition now would take actually getting an entitling explanation.

Nevertheless, the explanation that they get could be from anyone that they treat as a source of the belief. Explicitly securing mutual recognition after an immediate failure should just be taken as following from actually receiving an explanation for the testimony and treating it as an explanation for it. It might be asynchronous, and probably often is, particularly when it occurs through an intermediary: the recipient and source might not recognize each other at the same time because the testimony isn't given at the same time as it is received. It is also not a condition of mutual recognition that the rational source has to be identified as any particular individual, or an individual at all. Mutual recognition can be between a human and anything that functions *like* a rational source that provides explanations. This may count computers in the future, but what I have in mind is institutions. Outside of computers, we might receive a piece of anonymous testimony that we can treat as being from an institution. With computers, a piece of software might be developed by a large number of programmers, directed by a large number of managers,

etc. I take it that what we do not achieve mutual recognition directly with computers today, even when we personify them as “saying” their generated testimony, but rather with a source who set them up to say it. We often don’t know who exactly is behind this testimony. But as long as it provides what looks like beliefs, and they come with explanations, we can treat the intermediary as having a source behind it and potentially achieve mutual recognition with that source.

Moreover, as long as we get what look like explanations from it, we can also place trust in it (hence we often talk about placing trust in institutions). We might establish mutual recognition and trust with a multiplicity of individuals or institutions through one intermediary if the beliefs we get through it seem to come from different sources. Using different software and websites, for example, would seem to involve engaging with multiple different rational sources that are nonetheless each treated as unified rational sources. This allows my account of mediated testimony to apply to a much broader range of cases than would otherwise be possible.⁹

2.6. Conclusion: Securing Explanations

In this concluding section, I will address a couple of possible concerns with this account. First, I will discuss how this account only tells us the conditions of prima facie entitlement to testimony, but not when we are actually aware of fulfilling these conditions and being entitled. Second, I will discuss how there are many cases where prima facie entitlement might seem to be irrelevant because it isn’t possible or is overridden. I don’t think that these are concerns for my account because strictly speaking they fall outside of an account of prima facie entitlement to testimony. My account nevertheless suggests, as a way of dealing with both of them, that testimonial contexts should be made conducive to sharing and receiving more explanations.

First, although awareness of when we are prima facie entitled to testimony does not necessarily follow from fulfilling the conditions for prima facie entitlement that I have argued for, I don’t think that this is a concern unique to my account. What I mean by not being aware of when we are prima facie entitled is that we might not know when we recognize or are recognized by a source. For fulfillment of the conditions on prima facie entitlement according to PEP, we

⁹ This paragraph is inspired by Burge 1998’s (329) discussion of how we may be entitled to the testimony of something that resembles a rational source. Cf. Reeves and Nass 1996 (181–89) for a discussion based in the psychology literature of how we do not normally treat the source behind a computer as a programmer. They emphasize that we treat computers as having a variety of different sources behind them. But on my account, nothing prevents us from potentially recognizing any of them as a rational source.

have to have some understanding and treat some testimony as open to further explanation. We might take testimony without knowing whether we have understanding or treat it as open to further explanation. Even if we seem to fulfill the stronger conditions on our prima facie entitlement given in AEP by actually receiving an explanation for the testimony and treating it as an explanation for it, an explanation for the wrong belief or even a nonexplanation could be mistaken for an entitling one. This might be a particular concern with computer testimony, with which we might mistake for an explanation a calculation that cannot be used as a reason in an inference having the testimony as its conclusion. But all of these are kinds of mistakes that could be made with any reasoning. Discovering whether we are entitled is a separate issue from becoming entitled and often has to happen through additional reflection or belief sharing that shores up more explanations that could act as entitlements. One way to try to deal with this issue is to make testimonial contexts more conducive to sharing and receiving explanations. Intermediaries sometimes present special issues with this. If the intermediary is a designed human artifact like software, then there is a question of how to design them to this end that I will touch on in section 3.

Second, although there are, as I previously mentioned, many cases where prima facie entitlement isn't possible, or where prima facie entitlement is overridden, my account of prima facie entitlement is still important because it says what is needed for our autonomy to reflect to be respected. What these principles point to is how testimonial contexts can be guided toward securing prima facie entitlements. Testimony mediated by a computer that we don't expect could be further explained could, for example, overcome an immediate failure of mutual recognition if the software were programmed to give explanations. For computer-assisted mathematical proofs (such as for the four-color theorem), the testimony of the proof software would probably not be open to further explanation or not treated as open to further explanation by users who do not know how the proof software works, so they would require an actual explanation for prima facie entitlement. This explanation might consist in an explanation of how the proof software was programmed and what the software is doing in performing the proof. A programmer of the software would seem to already be prima facie entitled to the software's testimony on these grounds, as long as the software was programmed in a reasoned way rather than by random guessing. But other users would likely require an explanation from the programmer or another person who understands the software. The best way to ensure that such an explanation is

available when users receive the testimony is to design the software to provide such an explanation.

I think the way this example goes applies broadly to cases of generated testimony. I take it, for example, that we are more often *prima facie* entitled to the testimony of a search engine through the rational sources that we find behind its individual results than we are through being able to expect further explanation of how the search engine itself was programmed to give us these results. I think this calls into question our *prima facie* entitlement to any testimony that relies entirely on the search engine, such as its autosuggestions and other content that it generates. AEP points to the fact that a search engine could be designed to entitle us more often at the level of the search engine by giving some explanation of how the search results were reached. AEP can, therefore, be used as a guide in the design of software and websites to ensure that there is mutual recognition and hence *prima facie* entitlement to the beliefs generated by them. The lesson to be learned for computer testimony from my account of *prima facie* entitlement is that there is at least a risk that users will sometimes fail to treat testimony as open to further explanation. It is better to explicitly secure mutual recognition by designing software that can fulfill the stronger condition on *prima facie* entitlement given in AEP by actually providing explanations.

As stated in AEP, we not only have to receive an explanation for some testimony but also have to treat it as an explanation for the testimony to be *prima facie* entitled to it. This goes back to the issue of our awareness of entitlement. But the way to deal with this issue, I suggested, was to make testimonial contexts more conducive to sharing and receiving explanations. The actual exchange of explanations will make it more likely that we are aware of our entitlements at the same time as it ensures that our autonomy to reflect is respected.

3. Trust and Computer Testimony

3.1. Introduction: Trust and Social Norms

In section 3, I will situate trust in my account of belief sharing practices and apply it to examples from computer use in particular. Trust is important for these examples because it explains two possible consequences: either not trusting, which prevents belief sharing altogether, or misplacing trust, which threatens the one-sided enforcement of social norms. In this section, I define *prima facie* entitlement to trust a source as following from mutual recognition.

Interesting cases for our awareness of when we should not trust a source are when the status of mutual recognition with a given source changes from an immediate success to an immediate failure. Consider some of the examples I originally gave when outlining and contrasting each kind of testimony (section 2.3). We might initially think that a phone call or message is from a human only to find out that it was automated or generated. We might then stop recognizing a rational source behind the testimony. But in these circumstances, we are typically aware—*because* of the change in status of mutual recognition with the source from immediate success to immediate failure—that we should not trust the source unless we can actually get further explanation for the belief. In such cases, we are typically protected from deferring to the testimony without thinking of it as open to further explanation, but we are unlikely to trust. Such limit cases for mutual recognition help point to the ideal where, instead of awareness that we should not trust, we are *prima facie* entitled to trust a source.

My account of how we are *prima facie* entitled to testimony when our autonomy to reflect is respected can be simply extended to include a *prima facie* entitlement to trust. I think that we have a *prima facie* entitlement to trust under these circumstances because we can be assured that the beliefs we mutually accept are not imposed on us. Since the entitlement guaranteed by mutual recognition is only *prima facie*, we can always disagree with the source of testimony. Nevertheless, we are *prima facie* entitled to trust in these circumstances because when we do agree, we can be assured that the agreement is not coerced. In this way, trust provides the foundation for establishing mutually acceptable norms for our thought and behavior, while always leaving room to question these norms.

After expanding, in the following section, on how having the autonomy to reflect when we are *prima facie* entitled to trust means that we can lose this trust itself by reflection, I will show why we fail to be entitled to trust with so much contemporary computer use, which either leads to a lack of trust or the threat of social norms being one-sidedly enforced on us. But what comes after will be dedicated to showing that there is nothing inherent about intermediaries that should prevent us from being able to develop trust through them. They just present special challenges that can be dealt with, I will argue, by ensuring that they are designed for, and backed by, communities in which mutual recognition could already succeed. Success or failure of mutual recognition, I will conclude, is historically situated.

3.2. *Trust and the Autonomy to Reflect*

We are prima facie entitled to trust, like when we are prima facie entitled to testimony, when there is mutual recognition that guarantees our autonomy to reflect is respected. In this section, I will show how this means that the possibility of losing this trust itself remains open if we were prima facie entitled to it. If, on the other hand, we are not prima facie entitled to trust a source but do so anyway, then our ability to retract our trust is threatened.

If we are prima facie entitled to trust a source, then, I argue, the very same reflection involved in having second thoughts may also make us lose our trust. It can happen specifically when we reflect on what we treat as someone's beliefs and implicitly or explicitly determine whether there are any incompatibilities. For example, if some AI software explained that we were denied a loan for explanation x and we determine there was explanation y , implicit but unexpressed, and incompatible with explanation x , then we would lose trust. (Explanation x might be that we were denied the loan solely for having bad credit; explanation y might be that we were denied the loan because of our race.) We could discover explanation y , like any of our beliefs that we use in having second thoughts, through our own reflection or through additional testimony, even the testimony of the source of the original belief (in that case, we might have to decide whether their confession is enough to override the initial withholding). In either case, we can lose our trust in the source by reflecting on explanation x in light of explanation y and determining that they were incompatible. Losing trust in a source can just follow from having second thoughts when the reflection is on what we treat as someone else's beliefs.

This account fits with the overall picture that I have given where the incompatibility of beliefs is primary. I have argued that using beliefs in consequence relations involves having some understanding of their incompatibility with other beliefs. It is by reflecting on this incompatibility content that we can lose trust. Often, the relevant compatible or incompatible beliefs are related, or intended to be related, as explanandum to explanans, or as consequence to antecedent. In the example that I gave, reflecting on the incompatibility of the beliefs allowed us to call into question a consequence relation. We lost trust because a source gave some testimony that we were prima facie entitled to and we treated them as having an explanation that is actually incompatible with the testimony; it is an explanation for the wrong belief.

Understood in this way, trust simply follows from, and is lost with, our prima facie entitlement to testimony. But it can be used to explain the consequences of when we are or are

not prima facie entitled to testimony. Trusting only when we achieve mutual recognition with a source protects us from social norms being one-sidedly enforced on us by that source. Our autonomy to reflect is respected, so it remains possible for us to question the source, up to retracting our trust in them. Trust that we place in a source when there is not mutual recognition, on the other hand, threatens our autonomy to reflect in a way that could keep us from being able to retract this trust itself. If there is not mutual recognition with a source, it threatens a lack of trust or the possibility that social norms may be one-sidedly enforced on us—two common outcomes with computer testimony, as I will discuss in the following sections.

3.3. Trust and Software Testimony

In this section, I will give examples of testimony generated by programming for which there may be a lack of prima facie entitlement to trust. I will show how this creates contexts in which we either do not trust, preventing beliefs from being shared, or where we misplace trust, threatening the one-sided enforcement of social norms. This may happen with testimony that is not open to further explanation or just not treated by a user as open to further explanation, which would threaten their autonomy to reflect if they did trust the source.

First, for computer testimony it is important to note that we can place trust directly in a source or through an intermediary. As I argued in section 2.5, if the intermediary is a designed human artifact like a computer, then we can treat the source for a given belief received through it as singular. We might also place or fail to place trust in different sources that we encounter through one intermediary, but each may nevertheless be treated as singular. This is what often seems to happen with computers, where we encounter different sources through different software and websites.

With contemporary computer use, however, there seems to be a widespread lack of prima facie entitlement to trust as a result of sources not recognizing users. This describes the sort of context of belief sharing that is likely to arise from a corporation hiding what data is being collected from users, or an algorithm whose workings are unknown to them determining what will be shown to them. We might receive beliefs about what data is being collected, or about what a social media news feed or search engine shows us, without any explanations or while treating them as not open to further explanation. Since many of these beliefs require, as part of their explanations, details such as how a website or some software functions, we might approach

these sources with the attitude of not treating their testimony as open to further explanation. Even if some of their testimony doesn't require these details as part of its explanation, the explanation for its explanation might. Whereas with other humans with whom we experience an immediate success of mutual recognition the conversation often remains open so that we are able to continue asking for explanations of explanations, the explanations of software can bottom out quickly. A prima facie entitlement for some testimony might be overturned as we quickly reach explanations that are not themselves open to further explanation or that we do not treat as open to further explanation. To illustrate this point, let me return to an example that I introduced in section 2.3. The Iowa caucuses in February 2020 used a smartphone app to record and report results. We might be prima facie entitled to testimony as to the results if we treat it as open to further explanation. But as we look into the explanations for it, we are likely to hit a limit where we begin to not be able to get further explanation for the explanations themselves or just treat the explanations themselves as not open to further explanation. This is likely to happen in this case due to an apparent combination of manual error and issues with the app that reported the results (Cohn et al. 2020 and Koebler, Cox, and Maiberg 2020). Being unable to treat these explanations themselves as open to further explanation could channel upward to make us lose our prima facie entitlement to the original testimony as to the results. If we misplace our trust in a source and accept their testimony when we have lost our prima facie entitlement to it, then we are open to the threat that social norms may be one-sidedly enforced on us.

An example from the moral testimony literature (Howell 2014) that shows testimony not being treated as open to further explanation will help to explicitly show how this can result in the one-sided enforcement of social norms. Howell intends this example to show how we can fail to be entitled to testimony while having understanding, but he also uses it to produce an account of entitlement to moral testimony altogether unrelated to understanding. My reply will show that the failure of the person in the example to be entitled is rather a failure to see the understanding that he has as usable, which directly threatens his autonomy to reflect. It is a limit case for my account of prima facie entitlement, because it shows how even someone who has understanding that appears to be more than sufficient for reflection up to having second thoughts can be vulnerable to deferring to testimony while his autonomy to reflect is not respected. Under these circumstances, he can misplace trust in a source that has the power to one-sidedly enforce social norms on him.

The example uses the idea of a hypothetical piece of computer software called Google Morals of which one could ask moral questions and always receive the correct answer. Since it will also take all the details of my account of entitlement and trust to fully show what is going on, it will give a thorough idea of how social norms may be enforced by contemporary computer software in general. Here is the example:

Sam the Schizoid Ethicist: Sam knows the ins and outs of moral theory. He's read all the stuff, has taken numerous courses in normative and metaethics, and has written not one, but two excellent Ph.D. Dissertations. He defended each of them against the most rigorous examiners who subsequently suggested publication. The problem is, despite his expertise, he is of two minds about ethics. One of his dissertations was titled "In Defense of Utilitarianism," and the other was titled, "Against Consequentialism: A Deontological Manifesto." For just about any moral issue, he can provide a comprehensive explanation and defense of the Utilitarian verdict as well as the Deontological verdict. This is fine, of course, except when the views disagree. Suppose he is confronting an issue on which they disagree—whether, say, to harvest Healthy Harold's organs to save five others. This is an important decision, so he asks Google Morals "Which is true, utilitarianism or deontology?" It says "utilitarianism" so he judges that he should harvest Harold's organs. (Howell 2014, 397–98)

Howell gives this example of someone having understanding in the limit to show that they can fail to be entitled to testimony despite having understanding. Yet the significance of the example lies in the fact that Sam no longer treats himself as capable of using his understanding to have second thoughts about his beliefs. It's as though he does not even treat the belief he receives on testimony as related to his belief in utilitarianism rooted in his intricately developed explanation for it, even though it is apparently the same belief. He doesn't seem to be capable of using his understanding of the testimony in the same way as he might have been able to before. He uses the testimony to make a decision, and he might go on to use it in making decisions in the future, but it seems as though it is no longer a belief that he thinks should remain open to reflection. This goes to show that one can take testimony for a belief that one has a lot of understanding of, clearly sufficient for reflecting on it, while nevertheless losing the autonomy to reflect. Then the belief taken on testimony is treated—for Sam, applied in the decision that he makes to harvest the organs—just as if one did not understand its content at all.

Google Morals provides a compelling context in which to defer to testimony that the software cannot provide any explanation of and that we may not treat as open to our own further explanation. The software is likely to attract people like Sam who display a willingness to give up their autonomy to reflect, which suggests that the programmers of Google Morals don't recognize their users as rational. They give their users beliefs in a way that allows them to be treated as not open to further explanation. We can see clearly with this example, too, how the failure of recognition is mutual: it is easy to approach this search engine almost as an oracle with wisdom of mysterious origins, rather than as outputting the beliefs of a rational source. Sam experiences an immediate failure of mutual recognition with it, and he fails to be *prima facie* entitled to its testimony under PEP. And as long as Google Morals is not programmed to give its own explanations, he won't be *prima facie* entitled to the testimony under AEP either. This outcome is likely with really existing contemporary computer software and websites whose testimony isn't open to further explanation or just not treated by users as open to further explanation. Software and websites designed to give testimony without offering explanations, or just to be approached by users as if their testimony had no further explanation, are not likely to give the users a *prima facie* entitlement to trust because they are not likely to achieve mutual recognition with those sources. It does depend on the user, however. Someone using Google Morals is not forced to accept its testimony. The example shows poor Sam trusting the source and accepting the testimony, but someone not so trusting might choose to reflect back on their own beliefs or consult another source. They can be protected from threats to their autonomy to reflect and become entitled to their beliefs in other contexts of belief sharing. The user dependency of mutual recognition is something that will be discussed further in the conclusion to section 3 and the paper as a whole.

Since this example is an instance of explicitly moral testimony, we can see how, if we trust the source, it could explicitly result in the one-sided enforcement of social norms. Google Morals could be programmed to give an imperative to the user desired by the makers of the search engine, and if they approach it with an unwarranted trusting attitude and in the mode of completing their practical reasoning, like Sam does, it can then directly influence their behavior. This could also happen with theoretical reasoning that results in beliefs that can become rules for our thought and action. A lot of the testimony that we get from software is not in the form of an imperative, but other beliefs that we take on can nevertheless become implicit rules for us in

theoretical or practical reasoning. Moreover, as I have previously mentioned, when we take on a belief without treating it as open to further explanation, our autonomy to reflect is threatened not just for the testimony, but also for all the beliefs involved in our understanding of it. Social norms might be one-sidedly enforced by the other beliefs involved in our understanding of the testimony rather than directly. The example of Google Morals just helps to show a very explicit case where social norms could easily and directly be imposed on someone.

Thus far, the way to avoid this problem might seem to be for someone like Sam to do the reflecting entirely on his own. The example seems to imply that Sam already has a more than sufficient knowledge of the utilitarian and deontological positions, since he is able to thoroughly defend both, so that any testimony other than a simple answer to his question of which one is true will be unhelpful to him. But, on the contrary, given just how undecided Sam appears to be between the two positions, I argue that it would almost seem as if he needed to get further explanations for adhering to one or the other in order to make up his mind. In Sam's case, it's likely that a *prima facie* entitlement to some new testimony would be overridden by an incompatibility with his own thoroughly worked out beliefs, but then, on my account, he will nevertheless find himself in a context where he can get further explanation.

A common problem with software that serves as an intermediary, as I have argued, is that we often do not treat its testimony as open to further explanation. But software can be designed to offer explanations, unlike Google Morals. They present special challenges that have to be met, but there is nothing inherent about intermediaries that should prevent us from being able to become *prima facie* entitled to trust through them. It is more likely that testimony is not open or not treated as open to further explanation. If this doesn't occur right away with some testimony, it might quickly occur with an explanation of an explanation. This threatens either not trusting the source of the testimony or misplacing trust in it so that it can one-sidedly enforce social norms on us. But software just needs to be designed to offer explanations to avoid this. If this requires information about how, e.g., a search engine reached its results, then it might also have to be designed so that it is easier to learn some of the technical aspects of how it works. I will discuss more specifically how software should be made for particular purposes and particular communities after briefly considering problems of receiving testimony from another user through a computer, since they will point in a similar direction.

3.4. *Trust and User Testimony*

What I have not touched on in the previous section is how an immediate failure of mutual recognition with computer use does not occur just with testimony produced by some programmed procedures. It may also occur when communicating with another user and may be more common with computers because mutual recognition is harder to achieve insofar as we depend on certain bodily, vocal, and other indications that testimony is open to further explanation. I have focused on the examples of generated testimony in this paper because it is more straightforward to apply my account to them, but in this section, I will suggest the broader applicability of this account to beliefs shared between users.

Lack of recognition does not occur merely when we think that someone holds a particular belief without reason. Rather, it occurs when we think that they are incapable of committing to or entitling themselves to any beliefs at all, in virtue of who they are. So a failure of mutual recognition with another user may follow when we do not think that we are actually communicating with another user or when they think this of us, or when we just treat each other as such. We or they might think that the communication is with a “bot,” literal or figurative—a piece of computer software or someone whom we think is merely repeating some received testimony that they have no explanation for.¹⁰ We can’t have a prima facie entitlement to trust in these circumstances. Like when it occurs with testimony from a program, an explanation has to be produced to secure mutual recognition, except here, rather than coming from the program, it might come from the one that was not initially treated as rational. Since it seems to be a common problem with communication through computers, software could be designed to help facilitate this, but as failures of mutual recognition also occur outside of computer use, it speaks to wider problems that cannot all be addressed by revisions to technology.

Whether it is from a program or another user, computer testimony has been shown to raise problems that, I argue, point to the fact that software should be designed for particular purposes and particular communities. Software can be more easily designed to be learned by its users if it is intended only for a particular community, and difficulties with explaining generated testimony are easier to grapple with if software only gives a limited range of relevant testimony. If it is designed for, and backed by, communities in which there could be successful mutual

¹⁰ This is, in fact, the equivocal way in which the term “bot” has recently been used as a pejorative online. See Kazemi 2019.

recognition already, mutual recognition is likely to succeed for the beliefs shared between users as well. In my conclusion to section 3, I will begin to respond to these concerns by developing a notion of intermediaries that can serve their communities as what I will call *logical tools*.

3.5. Conclusion: Logical Tools for Representing Explanations

I have not yet discussed in detail, and it will largely be outside the scope of this paper to do so, specific recommendations for implementations of this theory in the design of computer software and hardware, in particular how a computer might best represent beliefs to make explanations accessible to testimony and reflection, or how software should effectively notify users of their right to an explanation. These are issues in the purview of semiotics and human-computer interaction that should be subjected to empirical study, on the basis of an account of what is required to achieve entitlement and trust. But in this section, I will briefly allude to how formal languages can provide a clue in this study. These are languages that generally have to be learned but are intended to be intelligible to anyone who learns them. I will argue that the transparency of an intermediary like a computer on explanations, which can secure mutual recognition through them, will follow from the extent to which it is able to represent both the inferences that we have made with beliefs and the further inferences that we can make with them as we continue to work on them in our belief sharing practices and reflection.

I will draw on Macbeth 2014's discussion of the Fregean senses of a content as making possible different inferential transitions.¹¹ I take her interpretation of sense to be the representational counterpart to my idea of the understanding that we have of a belief. Understanding refers to how we can use a belief in inferences; we can do this only by grasping a certain sense of the belief. But a sense is not just something subjective that we grasp, it can also be represented. So the question of how an intermediary can be transparent on our explanations can be translated into the question of how an intermediary might best represent our beliefs with senses that show the inferences that we have made with them and further inferences that we can make with them. This is no easy question, since different individuals have different understandings of the same beliefs. Even someone who is able to use a belief in the same

¹¹ A clear and detailed example is given in Macbeth 2014, 289–91, showing how different senses make working through a proof of algebraic equations possible. Cf. Shin 2011, who does not put it in terms of Fregean senses but nevertheless makes a similar point with regard to Peirce's existential graphs.

inferences as someone else might still have different implicit beliefs that are involved in the understanding that enables them to make those inferences. But I argue that we already have tools that are intended to help navigate these differences and continue to work on the original understanding we have of our beliefs. This is exactly what some formal languages are meant to do. They have to be learned, but once they are, they can be used to capture and refine our natural-language reasoning. For instance, they could be used to better display the logical equivalence of two different senses by prescribing inference rules between them that are, furthermore, visually perspicuous in the notation.¹² Learning a language that is mutually intelligible by anyone who has learned it makes it easier to share these different senses and convey similar understanding of beliefs.

I would like to generalize this idea beyond what we think of as formal languages. On what Brandom 2000 calls the expressivist view of logic, where logic's purpose is to make more of our reasons explicit, I think that any tool that allowed us to capture and continue to work on our understanding of beliefs through a mutually intelligible language could be called a *logical tool*. One could envision a programming language that allowed programmers to convey senses to users that were not significantly different from what the programmer worked with, or at least closed the gap by showing the equivalences.¹³ The resulting software would likely blur the line between programmers and users, which is a sensible ideal to strive for given the concerns that I have identified with entitlement to testimony generated by programming. This might require users becoming more literate in some of the technical aspects of how the software works, but I think this may just be what is necessary for prima facie entitlement in some cases. It is the same as the prerequisite of learning how to use a formal language for prima facie entitlement to the beliefs communicated through it.

There is nothing inherent about intermediaries that means that they have to prevent us from becoming entitled to testimony or trusting others through them. Any context will, like a game, set some of the rules according to which we share and reflect on our beliefs, and intermediaries are no different. If the intended users can learn how to use it, and it accurately codifies our understanding and allows us to continue working on it, an intermediary can further

¹² Schlimm 2018 describes how Frege's Begriffsschrift, for example, allows easy reading of certain combinations of simple symbols for conditionals and negation as complex symbols for conjunction and disjunction.

¹³ Reeves and Nass 1996 provide design recommendations based on empirical studies of how we typically treat computers similarly to other humans that might help with closing this gap.

open up the understanding to being communicated accurately to others and extended or revised by them. If there is mutual recognition through this process, then the intermediary can help us to place trust in each other, and by *prima facie* entitling us to beliefs, it can protect our autonomy to reflect on them. An intermediary could even be designed to give us distance from our beliefs that helps us to assume this reflective stance toward them. As I will argue in my concluding section, intermediaries that are designed for, and backed by, communities in which there could be successful mutual recognition are likely to be able to serve as logical tools for these communities.

4. Conclusion: The Historicity of Recognition

I hope my account throughout this paper has shown why it is important to go back into a pragmatic metavocabulary to discover what we mean by what we do. Understanding entitlement as instituted by social practices helps us to see how it is a social status. Whether mutual recognition as I have described it explains how entitlement is instituted exactly rightly is less important than the fact that entitlement to testimony and trust is a social status. Epistemology that does not analyze it as social just takes it as given and can't explore solutions, rooted in the practices that institute it, to the problem of one-sided enforcement of social norms that can result from mistakenly taking ourselves to have this status. In this concluding section, I will explore how the success or failure of a practice like mutual recognition that institutes our *prima facie* entitlement through an intermediary like a computer depends on the historical context of who recognizes and is recognized by whom. Applying my analysis of entitlement as a social status in this way will make it possible to identify existing problems and potential solutions.

A lack of mutual recognition can explain how, with contemporary computer use, we are often not *prima facie* entitled to accept beliefs or to trust because the sources and users do not recognize one another as rational. It may even be argued that the military, corporate, and other contemporary institutional sources behind the development of existing digital and network infrastructure do not in general achieve mutual recognition with some people, which calls into question which users they might respect the autonomy of through software. To take one contemporary example, the US military is funding research on explainable AI. The project managers admit that they are focusing on providing explanations for software used by the Department of Defense in intelligence analysis and war (Gunning and Aha 2019, 45–46). These

explanations are likely to suffice for the users that the source of the testimony has in mind, who need to be able to trust the source and be as assured as possible of their entitlement to its testimony for the purposes of their tasks. But war and intelligence analysis are activities that specifically withhold information from others. The interests of the source are served best by exposing actual explanations to some and hiding them from others. Software is always developed for some purpose and in the context of particular communities, from which some people will always be excluded from recognition. This is a concern I also identified in the introduction to this paper about the software used in criminal justice contexts. To recall, the software, COMPAS, uses the answers to a questionnaire that a defendant completes when they are booked in jail to generate a risk assessment score that is used to make decisions about the defendant. COMPAS seems to be designed to outsource, for time and cost efficiency's sake, at least some of the exchange of beliefs about a defendant's likelihood to commit further crimes. The intervention of this intermediary is likely to facilitate its users to trust its testimony while dismissing or even not receiving the direct testimony of defendants or other sources representing them who could provide further explanation. It has the consequence, in other words, of facilitating a context of belief sharing in which some but not others are recognized.

The ease with which computers might be set up to fail at mutual recognition with some users might even go back to their beginnings in Boolean logic. It might be a fraught task to identify what the social and material origins of Boolean logic exactly are, though one hypothesis that could at least explain the peculiar retrospective utility of Boolean logic for the invention of computers would be that there were already scientific and technological norms that went to shape this logic. These norms may have been such that the focus for designing a logic was not on creating a tool to help with sharing beliefs, but efficiency of calculation.¹⁴ Boolean logic's well-formed expressions notably are not always interpretable as *meaning* something. Getting from

¹⁴ Boole thought of logic as reducible to mathematical expressions:

On the principle of a true classification, we ought no longer to associate Logic and Metaphysics, but Logic and Mathematics. Should any one after what has been said, entertain a doubt upon this point, I must refer him to the evidence which will be afforded in the following Essay. He will there see Logic resting like Geometry upon axiomatic truths, and its theorems constructed upon that general doctrine of symbols, which constitutes the foundation of the recognised Analysis. (Boole 1847, 12)

That Logic, as a science, is susceptible of very wide applications is admitted; but it is equally certain that its ultimate forms and processes are mathematical. (Boole 1854, 8)

This suggests that mathematics does not itself have the form of reasoning for him. The converse idea that logic could provide a foundation for mathematics is a view that would only arise later and is more compatible with logic being a tool to help with sharing beliefs.

inputs to outputs involves performing calculations that do not have any sense that may be translated into language, and so seem to follow a procedure that involves a way of analyzing language from which much of our reasoning escapes. Boolean logic acts as a calculus ratiocinator, a tool for calculating to results, as opposed to a *lingua characteristic*, a language that captures how the reasoning was actually done, or how we understand beliefs. Being a *characteristica* is not mutually exclusive with being a calculus. In fact, on the view of logical tools that I have been expounding, a good *characteristica* would be a good calculus. If a logical tool helps us to capture our understanding and continue to work on it, then it is also good at calculating to results. But if Boolean logic well describes the material functioning of computers, then it may be that computers, as they currently exist at any rate, face limits to being good logical tools that are transparent on our belief sharing practices and reflection. If this helps them to give testimony in a way that does not *prima facie* entitle us to the testimony, then they may be easily used to intervene in our reasoning in a way that does not respect our autonomy to reflect.

This is not to say, however, that the solution lies in any potential future technology that could universally tend toward a successful practice of mutual recognition. It is unrealistic to expect any technology to play this role, just as it is unrealistic to expect that there could be any universal *lingua characteristic* for our reasoning. Logical tools are just tools; they have to be designed for, and backed by, contexts in which mutual recognition could already succeed. Such contexts, I will argue, are historically situated. A practice like mutual recognition allows us to integrate new beliefs while ensuring we still have the autonomy to reflect on them, up to being able to have second thoughts about them. Communication might even help us to engage in such reflection by making us attend to our beliefs in different ways. But if achieving mutual recognition ensures the autonomy to reflect, this does not have to suggest that there is any one ideal knowledge community toward which belief formation tends to converge, as Peirce 1868 has argued:

And what do we mean by the real? It is a conception which we must first have had when we discovered that there was an unreal, an illusion; that is, when we first corrected ourselves. Now the distinction for which alone this fact logically called, was between an *ens* relative to private inward determinations, to the negations belonging to idiosyncrasy, and an *ens* such as would stand in the long run. The real, then, is that which, sooner or later, information and reasoning would finally result in, and which is therefore

independent of the vagaries of me and you. Thus, the very origin of the conception of reality shows that this conception essentially involves the notion of a COMMUNITY, without definite limits, and capable of a definite increase of knowledge. (52)

Peirce thinks that our fundamental ability to correct errors must involve tending toward such a community. But I suggest there is nothing about having second thoughts about our beliefs that requires one such community. If he is wrong about this, belief formation is not on that account merely individualistic. Our beliefs are normative only by having some social status conferred on them, as I have argued, and as I think Peirce is alluding to here. But I have also shown that a practice like mutual recognition that institutes our entitlements can and does fail.

A better picture, I argue, is of a game that is by turns competitive and cooperative. Communities are formed around consistent mutual recognition, which allows us to find out what we agree on at the same time as it ensures respect for our autonomy to reflect. They will also compete with other communities with whom they disagree more deeply or are not even recognized by. This can account for both the conflicts that occur between players in the game, as well as the plethora of things about which they tend to agree, especially if they share the same culture, economic class, race, gender, or other identity features. It can also account for how a lack of mutual recognition is not in all cases negative, but may even be applied strategically by a community within which mutual recognition occurs but which requires, for the maintenance of the autonomy of its members, that others be excluded from such recognition. I don't mean by this fascist groups, which we could define as not only refusing to recognize people who do not belong to the group, but also as failing to maintain the autonomy of their own members because they do not facilitate a practice of mutual recognition within the group itself. I intend rather to point to communities threatened by the interests of those opposing their autonomy.

Logical tools should be developed to support such communities. The intermediaries in our social practices of belief sharing affect what we can do with our beliefs, just as the rules of certain games can make it possible or impossible, or more or less likely, to reflect and act in different ways and to compete against and cooperate with others. If these intermediaries are made by sources that do not in general achieve mutual recognition with the intended users, it is not likely that they will be able to serve them. Resistant use could be made of existing intermediaries, but there also has to be resistant design of intermediaries specifically for such communities. If an intermediary is a logical tool, then it should enable belief sharing that

respects our autonomy to reflect, to make explicit more of what those of us in the community agree on, or even to find out who is on our side, in the sense of being willing to fight for our fundamental beliefs or even our recognition. It should also enable us to discover and strategize against those who do not respect our autonomy rather than obfuscating the failures of mutual recognition that we experience with them. It should support us in taking action against them. Logical tools are just tools. But as long as we rely on them sometimes, they cannot come from sources that threaten our autonomy to reflect and hence to act without our use of the tools itself threatening our autonomy to reflect and to act. They can support us by ensuring respect for this autonomy.

References

- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. "Machine Bias: There's Software Used across the Country to Predict Future Criminals and It's Biased against Blacks." *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Boole, George. 1847. *The Mathematical Analysis of Logic, Being an Essay towards a Calculus of Deductive Reasoning*. Project Gutenberg. <http://www.gutenberg.org/files/36884/36884-pdf.pdf>.
- . 1854. *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*. Project Gutenberg. <http://www.gutenberg.org/files/15114/15114-pdf.pdf>.
- Brandom, Robert B. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.
- . 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge: Harvard University Press.
- . 2008. *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press.
- . 2019. *A Spirit of Trust: A Reading of Hegel's Phenomenology*. Cambridge: Harvard University Press.

- Burge, Tyler. 1993. "Content Preservation." In *Cognition through Understanding: Self-Knowledge, Interlocution, Reasoning, Reflection; Philosophical Essays, Volume 3*. 2013. Oxford: Oxford University Press.
- . 1998. "Computer Proof, Apriori Knowledge, and Other Minds." In *Cognition through Understanding: Self-Knowledge, Interlocution, Reasoning, Reflection; Philosophical Essays, Volume 3*. 2013. Oxford: Oxford University Press.
- . 2009. "Primitive Agency and Natural Norms." In *Philosophy and Phenomenological Research* 79 (2):251–78. <https://doi.org/10.1111/j.1933-1592.2009.00278.x>.
- Castro, Clinton. 2019. "What's Wrong with Machine Bias." *Ergo* 6 (15):405–426. <https://doi.org/10.3998/ergo.12405314.0006.015>.
- Cohn, Nate, Josh Katz, Denise Lu, Charlie Smart, Ben Smithgall, and Andrew Fischer. 2020. "Iowa Caucus Results Riddled with Errors and Inconsistencies." *The New York Times*. Feb. 6, 2020. <https://www.nytimes.com/2020/02/06/upshot/iowa-caucuses-errors-results.html>.
- Gunning, David and David W. Aha. 2019. "DARPA's Explainable Artificial Intelligence Program." *AI Magazine* 40 (2):44–58. <https://search.proquest.com/docview/2258093718>.
- Hills, Alison. 2015. "Understanding Why." *Noûs* 49 (2):661–88. <https://doi.org/10.1111/nous.12092>.
- Howell, Robert J. 2014. "Google Morals, Virtue, and the Asymmetry of Deference." *Noûs* 48 (3):389–415. <https://doi.org/10.1111/j.1468-0068.2012.00873.x>.
- Kazemi, Darius. 2019. "The Bot Scare." <https://tinysubversions.com/notes/the-bot-scare/>.
- Koebler, Jason, Joseph Cox, and Emanuel Maiberg. 2020. "An 'Off-the-Shelf, Skeleton Project': Experts Analyze the App that Broke Iowa." *Vice*. Feb. 5, 2020. https://www.vice.com/en_us/article/3a8ajj/an-off-the-shelf-skeleton-project-experts-analyze-the-app-that-broke-iowa.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.
- Macbeth, Danielle. 2014. *Realizing Reason: A Narrative of Truth and Knowing*. Oxford: Oxford University Press.
- Peirce, Charles Sanders. 1868. "Some Consequences of Four Incapacities." In *The Essential Peirce*, vol. 1. Edited by Nathan Houser and Christian Kloesel. Bloomington: Indiana University Press, 1992.

- . 1885. “On the Algebra of Logic: A Contribution to the Philosophy of Notation.” *American Journal of Mathematics* 7 (2):180–96. Johns Hopkins University Press. <http://www.jstor.org/stable/2369451>.
- Reeves, Byron and Clifford Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Stanford: CSLI Publications.
- Schlimm, Dirk. 2018. “On Frege's Begriffsschrift Notation for Propositional Logic: Design Principles and Trade-Offs.” *History and Philosophy of Logic* 39 (1):53–79. <https://doi.org/10.1080/01445340.2017.1317429>.
- Sellars, Wilfrid. 1956. “Empiricism and the Philosophy of Mind.” *Minnesota Studies in the Philosophy of Science* 1:253–329. <http://www.ditext.com/sellars/epm.html>.
- Shin, Sun-Joo. 2011. “Peirce’s Alpha Graphs and Propositional Languages.” *Semiotica* 186 (1):333–46. www.doi.org/10.1515/semi.2011.059.
- Teller, Paul. 1980. “Computer Proof.” *Journal of Philosophy* 77 (12):797–803. <https://doi.org/10.2307/2025805>.
- Tymoczko, Thomas. 1979. “The Four-Color Problem and Its Philosophical Significance.” *Journal of Philosophy* 76 (2):57–83. <https://doi.org/10.2307/2025976>.
- Wong, Kai-Yee. 2012 “Computer, Proof, and Testimony.” *Studies in Logic* 5 (1):50–67. <https://studiesinlogic.sysu.edu.cn/CN/article/downloadArticleFile.do?attachType=PDF&iid=49>.