

Special Issue VI  
2012

ISSN 1807-9792

# abstracta

*Linguagem, Mente & Ação*

**Précis of "Knowledge on Trust"**  
Paul Faulkner

**Faith in Others**  
Guy Longworth

**Knowledge on Affective Trust**  
Arnon Keren

**Can Trust itself Ground a Reason to Believe the Trusted?**  
Edward S. Hinchman

**The Trust Game and the Testimony Game**  
Katherine Hawley

**Testimony, Trust, and Social Norms**  
Peter J. Graham

**Replies**  
Paul Faulkner

**ABSTRACTA**  
*Linguagem, Mente e Ação*

ISSN 1807-9792

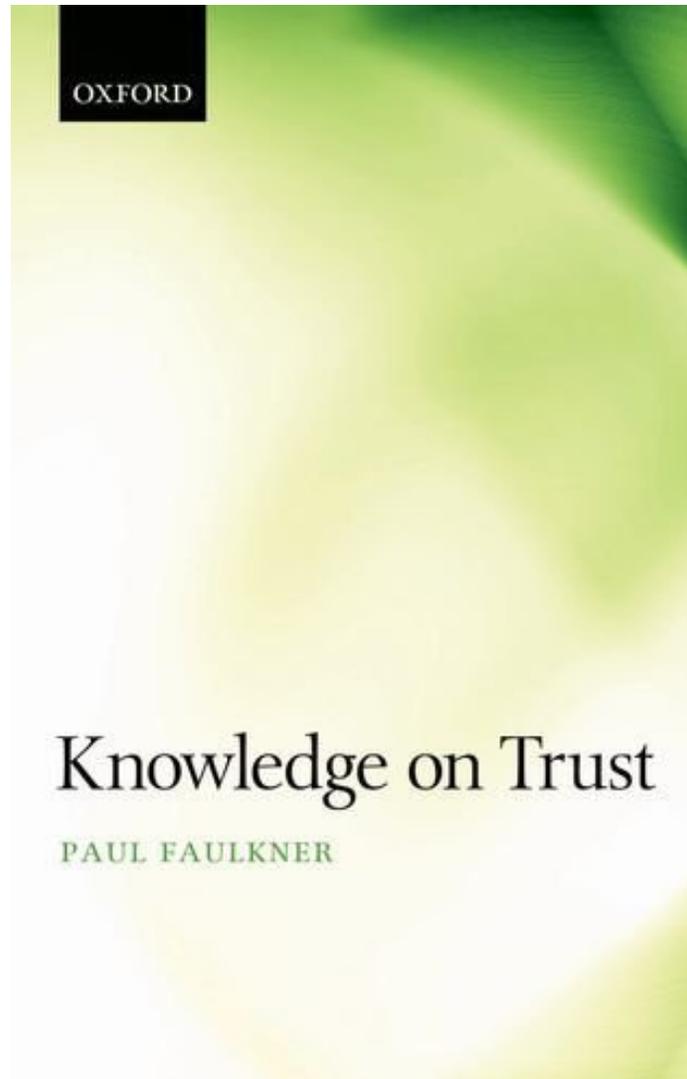
Special Issue VI  
2012

**Editors**

André Abath

Leonardo de Mello Ribeiro

A SYMPOSIUM ON PAUL FAULKNER'S  
*KNOWLEDGE ON TRUST*  
(OXFORD UNIVERSITY PRESS, 2011)



## TABLE OF CONTENTS

<b>Editorial</b>	<b>3</b>
<b>Précis of “Knowledge on Trust” (Oxford University Press, 2011)</b> Paul Faulkner ( <i>University of Sheffield</i> )	<b>4</b>
<b>Faith in Others</b> Guy Longworth ( <i>University of Warwick</i> )	<b>6</b>
<b>Knowledge on Affective Trust</b> Arnon Keren ( <i>University of Haifa</i> )	<b>33</b>
<b>Can Trust itself Ground a Reason to Believe the Trusted?</b> Edward S. Hinchman ( <i>University of Wisconsin-Milwaukee</i> )	<b>47</b>
<b>The Trust Game and the Testimony Game</b> Katherine Hawley ( <i>University of St Andrews</i> )	<b>84</b>
<b>Testimony, Trust, and Social Norms</b> Peter J. Graham ( <i>University of California, Riverside</i> )	<b>92</b>
<b>Replies</b> Paul Faulkner ( <i>University of Sheffield</i> )	<b>117</b>

**Editorial**

We are proud to bring to our readers the sixth special issue of ABSTRACTA, dedicated to *Knowledge on Trust* (Oxford University Press 2011), written by Paul Faulkner from the University of Sheffield.

In *Knowledge on Trust* Faulkner presents a theory of testimony, according to which one can obtain knowledge on the basis of testimony if and only if one has reasons to trust the speaker. Faulkner offers an account of a form of trust – which he calls affective trust – that can play the epistemic role of providing rational support for beliefs obtained via testimony. According Arnon Keren, one of the commentators in our symposium, Faulkner’s “explanation of how trust can be rational, and how beliefs formed through trust can be epistemically warranted, is highly original.” (p.34)

As a significant contribution to the growing literature on the epistemology of testimony, *Knowledge on Trust* deserves extended discussion, and that is what we tried to bring out with this symposium. We would like to thank, first of all, Paul Faulkner himself, for his time, attention and unfailing support towards this project. Secondly, we would like to thank the authors, who took their time to write substantial pieces on *Knowledge and Trust*: Guy Longworth, Arnon Keren, Edward S. Hinchman, Katherine Hawley and Peter J. Graham. Last, but not least, we would like to thank Oxford University Press for their support.

André Abath &  
Leonardo de Mello Ribeiro,  
EDITORS.

December, 2012.

**PRÉCIS OF “KNOWLEDGE ON TRUST”  
(OXFORD UNIVERSITY PRESS 2011)**

**Paul Faulkner**

The ambition of *Knowledge on Trust* is to offer a theory of testimony, or an account of how we get to know things on the basis of testimony. A very short outline of the position I develop might run as follows. One type of account – the *non-reductive theory* – proposes that facts about the rationality of speakers and what is involved in understanding underwrite a general entitlement to believe testimony. We then get to know things on the basis of testimony when the bit of testimony believed expresses knowledge. This account, I argue, is right to conceive of the epistemology of testimony in *transmissive terms*, but is wrong to propose a general entitlement to believe testimony. The problem is that giving testimony is a practical activity, so there are a multitude of explanations that can be given for any piece of testimony and a particular reason is needed for believing any bit of testimony expresses knowledge. This follows from, what I call, the *problem of cooperation*: in abstract the interests of speakers (having an influence) are orthogonal to the interests of audiences (being informed). This problem is resolvable in that we do, in fact, have ample reasons for believing testimony. Such a reason would be, for instance, the belief that the speaker had the one’s informative needs at heart, or was a reliable voice of truth, or had property X, Y or Z. The *reductive theory* of testimony then proposes that we get to know things on the basis of testimony just when and because our reasons for belief are knowledge supporting. This account, I argue, is right to demand that testimonial uptake be rationally supported, but is wrong to give it a reductive epistemology. Testimonial knowledge is transmitted and, as such, the epistemological role played by our reasons for believing testimony is just that of making it reasonable for us to believe that a bit of testimony is an expression of knowledge. Trust, I then argue, can play this epistemological role, and specifically trust in a thick normative sense, which I call *affective trust*. Affective trust is a second-personal notion in that it involves attitudes about another’s attitudes to oneself. Both reductive and non-

reductive theories then fail in not recognizing how testimony can be situated in and institute a testimonial relationship that is structured by trust and its presumptions. This fact about testimony is recognized by the *assurance theory*, and *Knowledge on Trust* offers some defence of this theory. The problem for the assurance theory is that it seems limited to face-to-face testimonial situations structured by trust; the problem is one of explanatory scope. *Knowledge on Trust* then offers a genealogical argument that takes this testimonial situation, and the problem of cooperation confronted in it, as the basic starting point and explains how our thinking about trust resolves this problem. So I argue that trust based reasons are not merely a possibility for us, and a possibility that is limited in various ways, but are fundamental to the institution of testimony.

**Paul Faulkner**

*University of Sheffield*

paul.faulkner@sheffield.ac.uk

## FAITH IN OTHERS<sup>1</sup>

Guy Longworth

### 1. Introduction

When someone tells us something we appear to face a choice: we can accept what they tell us, or not. Insofar as we face a genuine choice—insofar as it is up to us whether or not to accept what we are told—our deciding in one way or another might be guided by any of a variety of considerations. Perhaps it is important to us, for practical, prudential, or theoretical reasons, to take a stand as to whether or not something is so. In that case, we might choose to accept what we are told because that is a way of taking such a stand, rather than remaining neutral. More typically, we hope not only to take one or another stand, but to take a stand that is correct or amounts to knowledge. In that case, we accept what we are told because we hope that it is a way to come to believe correctly, or to know, where the latter is something that we desire or need. Alternatively, perhaps we recognize that our interlocutor is competent with respect to what they tell us, so that failing to accept what we are told would amount to treating them as insincere. (Here ‘recognize’ takes an epistemic, rather than honorific, reading: to recognize is to know, rather than to acknowledge an entitlement.) In that case, we might choose to accept what we are told because treating a person as insincere without good reason is in conflict with the demands of morality. Or perhaps we recognize that our interlocutor is sincere, so that failing to accept what they tell us would amount to treating them as incompetent. In that case, again, there may be moral reasons for accepting what we are told. And, of course, we may recognize neither the competence nor the sincerity of our interlocutor and yet acknowledge both on moral grounds.

---

<sup>1</sup> I'm grateful for discussion and advice to Gary Banham, Paul Faulkner, Hemdat Lerman, and Matthew Soteriou.

It seems, then, that there may be prudential or moral reasons—that is, practical reasons—for accepting what someone tells us. There may also be straightforwardly cognitive or epistemic—that is, theoretical—reasons. (The reason for the adverbial qualification will become clear shortly.) For instance, we may recognize that our interlocutor is both competent and sincere, perhaps by apprehending that they speak from knowledge. But the more typical situation is one in which we have at best indecisive grounds of that sort for believing that an interlocutor speaks from knowledge. In those more typical situations, where we might seem to lack straightforwardly cognitive or epistemic reasons for accepting what we are told, it seems that so accepting in spite of that lack is an exercise of *faith*. If the appearances here are veridical, then accepting what we are told is often, and perhaps predominantly, an act of faith. Yet many people now hold that accepting what we are told in such situations can be a way of coming to know. How, if at all, can an act of faith give rise to knowledge?

To give the question bite, consider the following line of argument.

1. In order to know, we must have cognitive or epistemic reasons—that is, theoretical reasons—that are decisive.
2. There are situations in which we reflectively take ourselves to be in a position to acquire knowledge by believing what we are told and yet the cognitive or epistemic reasons to which we have access independently of faith are not decisive.
3. The reasons to which we have access are either made accessible to us independently of faith or are made available by faith.
4. Faith is not itself a source of decisive cognitive or epistemic reasons; it is a source only of practical reasons.
5. Faith cannot furnish access to decisive cognitive or epistemic reasons. (The only reasons to which one has access in the situations operative in 2. are either indecisive or have their source in faith.)

Therefore:

6. The situations described in 2 are situations in which the cognitive or epistemic reasons to which we have access are not decisive.

7. In those situations, we cannot come to know by believing what we are told.

As noted, many people now hope to resist the conclusion in 7. That is, many people hope to defend the claim that we can know on the basis of accepting what we are told, even where the reasons to which we then have access, independently of faith, are not decisive. Disagreement within that group concerns two questions:

(Q1) Which of the operative premises, 1–5, are to be rejected?

(Q2) On what grounds are those premises to be rejected?

I'll focus here mainly on premises 4 and 5, considering some of the options for resisting those premises, and some of the challenges that those options entail. My discussion is a variation on some central themes in Paul Faulkner's important and insightful book, *Knowledge on Trust*. My aim is to outline some questions that remain open concerning the terrain that Faulkner seeks to map. First, I'll attempt to indicate some areas that remain to be explored (Sections 2 and 3). Second, I'll try to indicate some ways in which some of the paths that Faulkner argues to be impassable might yet be navigated (Sections 4 and 5). I'll conclude with some questions.

## 2. Faith

The claim at issue in this section is that faith is not itself a source of cognitive or epistemic reasons. (Premise 4 in our opening argument.) The claim has an impressive pedigree, with Kant as its most articulate proponent. Kant distinguishes faith from knowledge by appeal to the status of their respective grounds and, thence, the sources of those grounds.<sup>3</sup>

According to Kant, someone's grounds for taking something to be true may be *subjective*, *objective*, or a combination of both. Where one has sufficient *subjective* grounds for taking it to be true that *p*, the question whether *p* is closed *for one*: in taking it to be true that *p* one does not allow for the possibility that it's not true that *p*. However, one's

---

<sup>3</sup> See especially Kant 1781/1787: A820/B848–A831/B859, 1785. I've been helped in my understanding of Kant's views in this area by Stevenson 2003 and Chigwell 2007, although our readings differ at various points.

possession of subjective grounds for taking it to be true that *p* is compatible with one's lacking objective grounds. Where one has, in addition, sufficient *objective* grounds for taking it to be true that *p*, one has grounds that would determine *for anyone* operating in accord with theoretical reason that it is true that *p*: one has access to theoretical grounds that demonstrate truth. On the basis of his distinction between subjective and objective grounds, Kant distinguishes three categories of taking to be true. One *knows* that *p* where one takes it to be true that *p* on grounds that are both subjectively and objectively sufficient. One *opines* that *p* where one takes it to be true that *p* in the consciousness that both one's subjective and one's objective grounds are insufficient. Faith is then the middle category: one *has faith* that *p* where one takes it to be true that *p* on sufficient subjective grounds, in the consciousness that one's objective grounds do not furnish a theoretical demonstration of truth.

Suppose that one were rational—that is, that one were operating in accord with reason. In that case, wouldn't one's consciousness that one lacked sufficient objective grounds for taking it to be true that *p* undermine one's subjective grounds, by revealing to one that one's theoretical grounds leave open that it may not be true that *p*? And in that case, isn't rational faith impossible? Not according to Kant. For Kant allows that reason takes practical as well as theoretical form. Where one has faith, one's conviction is held in place by the demands of practical, rather than theoretical, reason. Those demands have their source in one's own will, rather than in factors that are impartially accessible to theoretical reason. One's rational grounds overall therefore can be impervious to the travails of one's merely objective grounds.

Kant holds that faith, when so understood, is possible only where three conditions are met. First, as noted, one's taking something to be true must be decided on practical, rather than theoretical, grounds. Second, its practical grounds must be unconditioned. Only in that circumstance can the grounds provided by practical reason be subjectively *sufficient*. That is, the only practical grounds that can serve rationally to *close* a question are grounds that are available independently of answers to further practical or theoretical questions. For Kant, that means that faith is restricted to what one is required to hold true in order to implement the demands of *categorical* morality. Third, and in effect a corollary of the

second condition, there must be no possibility that the question at issue might be decided theoretically. It must be impossible for the question to be answered theoretically in a way that conflicts with practical requirements, on pain of irreparable discord between the demands of practical and theoretical reason. And, for Kant, it must also be impossible for practical and theoretical reason to determine the same answer. For if that were possible, then grounds that were practically decisive would nonetheless leave open the possibility of discord, and so reveal the practical grounds to be, at best, conditional.

In addition to cases of faith proper, Kant allows for other cases of taking something to be true on practical grounds, where those grounds do not close the question at issue. For instance, Kant provides the following example of what he calls *pragmatic belief*:

The doctor must do something for a sick person who is in danger, but he does not know the illness. He looks to the symptoms, and judges, because he does not know of anything better, that it is consumption. His belief is merely contingent even in his own judgment; someone else might perhaps do better. (Kant, 1781/1787: A824/ B852.)

In this case, the doctor's taking it to be true that the patient has consumption is based upon theoretical grounds that are neither subjectively nor objectively sufficient. It is held in place by the practical demand that the doctor acts, and acts on relevant theoretical beliefs. But the doctor's practical grounds are doubly insufficient to decide what he should take to be true. First, they are insufficient in that the doctor's practical end—treating the patient—is contingent: the doctor might not have set that end for himself. If he had not, then he may have lacked practical grounds for taking it to be true that the patient has consumption. Second, they are insufficient in that, having set for himself that end, the doctor's practical grounds still leave open for him the question, whether the patient in fact has consumption.

Returning to premise 4, it would be natural to take the notion of faith that appears there to correspond, not with Kant's very demanding technical notion of faith, but rather with his notion of pragmatic belief. Where we take what we are told to be true, in the sorts of cases at issue here, our doing so is decided for us, in part, on practical grounds. We are aware that it is impossible for us decisively to close the question at issue on purely theoretical grounds. And yet our broadly practical ends—including, most minimally, the

contingent end of attaining an answer to the question at issue—incline us to adopt an answer to the question. Assume that our being told what we are provided us with some, albeit indecisive, grounds to adopt one rather than another answer—grounds that may themselves be theoretical or practical. In that case, our practical ends decide for us that we will adopt that answer, while leaving open that there may be further theoretical grounds that either support or undermine that answer. There would therefore be at least two reasons for denying that our taking what we are told to be true would amount to faith in the demanding Kantian sense. First, the operative practical ends would be contingent: we might have lacked the ends that induce us to adopt an answer to the question at issue. Second, the operative ends are anyway indecisive: adopting them leaves open that the question may be answered differently, or more effectively, on other grounds.

On the construal of faith as (Kantian) pragmatic belief, it would be over-determined that faith could not provide decisive theoretical grounds. For pragmatic belief is held in place by grounds that are practical, rather than theoretical; and—for the reasons outlined above—it is held in place only provisionally. On the assumption that knowledge must be based upon decisive theoretical grounds—premise 1 in our target argument—it would therefore be over-determined that faith cannot on its own sustain knowledge.

A response to that interim conclusion might take any of the following forms. First, one might accept the interim conclusion and so—assuming acceptance of the target argument's other elements—accept that taking what one is told to be true is itself a form of pragmatic belief rather than knowledge. Second, one might accept the interim conclusion but avoid accepting that taking what one is told to be true is itself a form of pragmatic belief by rejecting another element in the target argument (most plausibly, by seeking to reject premise 5, and attempting to argue that faith can provide access to reasons that are constituted independently of that faith). Third and fourth, one might reject the interim conclusion by accepting that one's grounds are partly practical, while denying that one's practical grounds are indecisive, giving rise to analogues of the consequences of response one and two with Kantian faith in place of pragmatic belief. Fifth and sixth, one might reject the interim conclusion by denying that the practicality of one's grounds is inconsistent with their also being, or giving rise to, theoretical grounds. Holding in place

the rest of Kant's taxonomic scheme, that response has two subdivisions, depending on whether the theoretical grounds provided by faith are themselves decisive or indecisive. Fifth, then, if it is held that faith can give rise to *indecisive* theoretical grounds, then the result will be that our taking what we are told to be true is, at best, the adoption of opinion. Sixth, if it is held that faith can give rise to *decisive* theoretical ground, then the result will be that our taking what we are told to be true may comprise our acquiring knowledge. Notice that only the sixth position sustains the view that faith can amount to knowledge.

Faulkner's view is that the form of faith involved in accepting what one is told—what he calls *trust*—is not itself a form of knowledge. This is so, in part, because he holds that reasons that are grounded in trust are indecisive. On his view, trust need not comprise decisive grounds, since its primary role in the justificational architecture of testimonial belief is not to provide grounds for those beliefs. Rather, its function is to ground one's acceptance of what one is told and so to provide one with access to reasons that are constituted independently of trust. Thus, Faulkner seeks to reject premise 5 by adopting a combination of the second and fifth options sketched in the last paragraph. As we'll see in the next section, those options bring in train a range of explanatory burdens due to the way in which they threaten to pull apart subjective and objective reasons.

### 3. Architecture

Suppose that the acceptance of what we are told, *per se*, is at best a form of faith or pragmatic belief. That is, suppose that we have been forced to adopt either the first, second, or fifth responses to Kant's position that were outlined at the end of the previous section. In that case, accepting what one is told does not alone amount to knowing that which one is told. And one's faith in one's interlocutor's competence and sincerity with respect to what they tell one does not alone amount to one's knowing that they are competent and sincere. For similar reasons, faith that a combination of one's interlocutor and their sources, or their sources' sources, etc., are competent and sincere does not alone amount to knowledge. On those suppositions, there will be a range of cases with the following five characteristics. First, they will be cases in which a speaker *S* tells an audience *A* something. Second, they will be cases in which *S* is in fact sincere and competent with respect to what they tell *A*:

their telling *A* what they do is an expression of their knowledge. Typically, we would take *A* to be in a position to come to know *p* in those circumstances, by accepting what they were told. Third, then, *A* can come thereby to know *p*. Fourth, *A* accepts what they are told on faith—that is, their acceptance of what *S* tells them is based only on faith in *S*'s competence and sincerity, on their having spoken from knowledge. Fifth, then, since *A*'s acceptance is based only on faith in *S*'s having spoken from knowledge, and since faith is not a form of knowledge, we have the following:

- (I) *A* knows *p*
- (II) *A* does not know that *S*'s telling *A* *p* is based on knowledge.

Many philosophers hold that one can acquire knowledge on the basis of accepting what one is told only if one's interlocutor meets conditions in addition to simply speaking the truth. One natural view is that, in the most normal cases, if one is to come to know *p* on the basis of accepting what one is told, one's interlocutor must know *p*. This view is subject to plausible (though indecisive) counterexamples. For example, someone who knows that smoking is carcinogenic—say, our family doctor—tells me that it is. I do not accept what they tell me, because I believe that they are simply propounding government propaganda. Nonetheless, I wish to discourage my daughter from smoking, so reliably transmit what I am told, on grounds that it will not be corrected by our family doctor. Plausibly, at least, if my daughter accepts what I tell her, she can thereby come to know that smoking is carcinogenic. If we take the case to be normal, then the case arguably presents a counterexample to the natural view.<sup>4</sup> According to a less demanding view, not threatened by the plausible counterexample, one's interlocutor's telling one what they do must be reliably based on knowledge, without that requiring that the knowledge is wholly in the interlocutor's possession. Since the details won't matter, and since Faulkner accepts a similar requirement, I'll assume that the less demanding view is correct.

---

<sup>4</sup> See e.g. Graham 2000.

According to the less demanding view, it's a necessary condition on someone's acquiring knowledge by accepting what they are told that what they are told is based on knowledge. Suppose that *A*'s reasons for accepting *p* are exhausted by reasons that *S* makes available by telling them *p*. In that case, we have (III):

(III) If *A* knows *p*, then *S*'s telling *A* *p* is based on knowledge.

Now, supposing (III) to be true, it's plausibly something that someone might know, at least when framed at the current level of detail. Let's suppose, then, that *A* knows (III). (The argument might be reconstructed on the weaker assumption that *A* merely believes (III), but it can be presented more simply if we make the stronger assumption that *A* knows (III).) So, we have (IV):

(IV) *A* knows that (if *A* knows *p*, then *S*'s telling *A* *p* is based on knowledge).

Now we would ordinarily expect a subject who is rational, and who possessed normal conceptual competence, to be able to progress beyond (I)–(IV) in at least two ways.

First, we would ordinarily expect a subject who knows something to be in a position to attain reflective knowledge to the effect that they know that thing. There is a range of cases where that expectation may not be met. For one example, self-knowledge might be blocked by idiosyncrasies in a subject's conceptual competence, for instance by their imposition of conditions on applications of the concept of knowledge that are too demanding. For another example, self-knowledge might be ruled out because the subject holds misleading beliefs, for instance to the effect that their first order belief is at best accidentally true. For a third example, self-knowledge might be foreclosed because the subject's first-order cognition is only just safe enough to count as knowledge so that correlative second-order cognition would not be.<sup>5</sup> However, although there plausibly are

---

<sup>5</sup> For instance, suppose that a subject's belief underwrites knowledge, because the belief is true in a range of nearby possible cases, but in many of those cases the belief is true but unsafe—for instance, because in too many cases near to those cases, the belief is false. So, although the first-order belief is safe enough to underwrite knowledge, it does not *safely* underwrite knowledge: there is significant danger that the subject

cases where one can know without being in a position to know that one knows, it is perfectly normal, at least with respect to ordinary cases of first-order knowledge, to be in a position to know that one knows. Hence, in many normal cases, one would expect *A* to be able to progress from (I) to (V):

(V) *A* knows that *A* knows *p*.

Second, we would ordinarily expect a normal subject who knows *p*, and knows that if *p*, then *q*, to be in a position competently to deduce *q* on the basis of those two pieces of knowledge. And it would be very natural to expect that, in normal cases, a competent deduction based on knowledge would result in knowledge. Again, there might be a range of more or less special cases in which those expectations are not met, but in normal cases, it would be natural to expect such a subject to be in a position to know *q*.<sup>6</sup> If *A* were in such a case, we would therefore expect them to be in a position to arrive, by competent deduction based on the knowledge ascribed in (IV) and (V), at the knowledge ascribed in (VI):

(VI) *A* knows that *S*'s telling *A p* is based on knowledge.

However, (VI) contradicts (II). Holding fixed our other assumptions, it appears that we have two *prima facie* unhappy options. First, we might reject (V), and claim that, in the type of case at issue—cases in which one knows only through faith-based acceptance of what one is told—one cannot know that one knows. A natural replacement position leaving intact as much as possible of the remaining structure, would then be that we can have only faith, and not knowledge, concerning our first-order knowledge. Second, we might reject (VI), and claim that, again in the type of case at issue, one cannot acquire the type of knowledge attributed in (VI) on the basis of competent deduction from things that one

---

might believe what they do without knowing it. Now suppose that the second-order belief—to the effect that the first-order belief underwrites knowledge—co-varies with possession of the first-order belief. In that case, the second-order belief may not be safe enough to underwrite knowledge, since there is significant danger of the subject having that second-order belief when, at first-order, they have only a true belief and not knowledge. See Williamson 2000, chapters 4 and 5.

<sup>6</sup> For discussion see e.g. Drestke 1970, Nozick 1981, Sherman and Harman 2011, Stine, 1976, Vogel 1990.

knows. Perhaps in such cases, the subject is in a position to see that something *must* be so, given what they know—or in order that what they take themselves to know can be so—but is not in a position to *know* that it is so. Again, a natural replacement position would be that competent deduction in such cases is a way of extending the reach only of faith, and not of knowledge.

A different style of response would involve denying that the standing of (II) is fixed once and for all by its genesis in faith. Instead, an attempt might be made to argue as follows. First, it is accepted that (II) holds at the point at which *A* first acquires knowledge from *S*. At that point, although *A* knows that which *S* tells them, *A* has only faith, and not knowledge, concerning the epistemic basis of the telling. However, second, it is argued that, having now acquired that knowledge, it is possible for *A* to use it in order to bootstrap their way into knowledge, rather than faith, concerning the epistemic pedigree of what they were told. Such a response might be motivated by reflection on the special nature of our knowledge of our own minds. For, on the one hand, it is apt to seem plausible that one can know things about one's own mind that depend upon connections with our environment. For example, it is apt to seem plausible that one can know that one knows that it's currently sunny, where one's knowing that depends on it's being sunny and on one's being appropriately sensitive to that fact. And, on the other hand, it is apt to seem plausible that one can have such knowledge prior to checking that the required environmental connections obtain. For example, it is apt to seem plausible that one can know that one knows that it's currently sunny without testing whether one is appropriately sensitive to its being so.

The viability of such a response appears to depend essentially on the partial independence of the acquired first-order knowledge from the faith on which its initial acquisition is based. That is, it requires that, although faith may play an essential role in instituting knowledge, its role cannot be that of a lemma in an argumentative justification for that which is known. For if it were, then *A*'s faith in the epistemic basis of what they were told would figure as an essential supporting element *throughout* *A*'s derivation of the knowledge ascribed in (VI). And that would make mysterious how the derivation could result in knowledge, rather than faith, concerning that epistemic basis. By contrast, if faith

functions as something more akin to an *enabling* condition for the acquisition of knowledge that is sustained independently of faith, then it is possible to at least make out the shape of a position able to sustain (V) and (VI).

If such a position can be made out, then we have a pleasing convergence with Faulkner's preferred view of the role of faith, or trust, in providing access to independently constituted reasons. However, it remains to be seen whether such a position is defensible. I would be interested to know more about Faulkner's own preferred response to the issue sketched in this section, and his reasons for preferring that response to other available responses.

#### **4. Responsibility**

The position to this point is this. It's plausible that faith in others figures essentially in many cases in which we take ourselves to acquire knowledge by accepting what we are told. However, large questions remain open. In particular, it remains an open question whether faith can comprise, or give rise to, theoretical reasons for taking something to be true. And it remains an open question what precisely the role of faith is in underpinning knowledge. However, in advance of more detailed examination of the various options for answering those questions, the following is plausible. With respect to the first question, it's plausible that faith cannot *alone* give rise to knowledge of what one is told, either directly, or via knowledge that the telling is appropriately based on knowledge. Insofar as it is possible to acquire knowledge by accepting what one is told, the underpinning of the knowledge must plausibly go beyond what is independently available to the faithful and make reference also to the epistemic pedigree of the telling. With respect to the second question, it's plausible that the role of faith is, not to provide lemmas in argumentative justifications for accepting what one is told, but rather to furnish an epistemic environment that is suitably hospitable to reasons that are constituted independently of one's faith.

Let's assume that the epistemic underpinnings of one's taking to be true what one is told are constituted independently of one's faith in the speaker. Whether those underpinnings can sustain knowledge is a matter of whether or not the telling is appropriately based on knowledge. When one is told something, one is either presented

with an opportunity to know, through accepting what one is told, or—because the telling is not appropriately based on knowledge—one is not. Let's call instances of the former type *good cases* and instances of the latter type *bad cases*. Thus far, we've been assuming that one lacks *decisive* theoretical grounds for taking it that one is in a good case—one in which acceptance will grant one knowledge—rather than a bad case. And that is, of course, consistent with one's having *indecisive* theoretical grounds for accepting what one is told. However, if one's grounds were entirely neutral—either because one lacked grounds for accepting or rejecting what one was told, or because, although one had grounds of both sorts, they were in equilibrium—one would lack any reason for accepting, rather than refusing to accept, what one was told. In that case, accepting what one was told, rather than remaining neutral about it, would evince doxastic irresponsibility.<sup>7</sup>

Now, we are treating faith-based acceptance of what one is told as serving, in effect, to enable one's belief to draw on the epistemic credentials of the telling, rather than as partly constituting that belief's epistemic underpinnings. It is not, therefore, entirely obvious that the doxastic irresponsibility of acceptance *must* have a negative effect on the epistemic standing of beliefs formed on its basis. However, although it is not obvious, it is plausible. Consider cases in which, despite the proper functioning of a resource, one has weighty evidence that the deliverances of that resource are *not* to be trusted. For instance, consider a case in which one knows that one's visual system is malfunctioning in such a way that it regularly seems to one visually as though there is movement at the periphery of one's visual field even though there is no such movement in one's environment. One seems to see such a movement and takes the seeming to be the upshot of the malfunction. In this particular case, one in fact glimpsed a cat leaving the room. Nonetheless, it's plausible that one is precluded from coming to know, on the basis of one's visual experience, that something moved. Generalizing, it's plausible that, even in cases wherein one's faith in an epistemic source would normally play only an enabling role, in allowing one to draw on reasons made available by the source, reasons for thinking that the resource is malfunctioning can preclude the source from playing that role. And it's plausible that the

---

<sup>7</sup> One way of thinking about issues of doxastic responsibility is as concerned with the appropriate connections between subjective and objective reasons.

same would hold with respect to cases of *genuine* neutrality—cases in which one really lacked *any* positive reason to accept the deliverances of an epistemic source.

It's therefore natural to wonder, with respect to cases of the sort we are considering, where we take ourselves to be in a position to acquire knowledge by accepting what we are told, whether any of them are cases in which our reasons for acceptance are in equilibrium, or even deficit. (More generally, we might consider whether any of them are cases in which we are sufficiently close to equilibrium that what would otherwise be an opportunity to acquire knowledge is screened off from us.) Faulkner considers, and rejects, a number of proposals according to which we have general, positive grounds for accepting what we are told, before offering in their place his own trust-based account. In the remainder, I want briefly to consider two of the accounts that Faulkner rejects, and raise some questions about his grounds for rejecting them.

One putative source of positive reasons for accepting what one is told would be what is presented to one, or made manifest to one, by a speaker who tells one something. On one such account—developed by John McDowell—a speaker, in telling one *p*, can thereby present one with whatever reasons underpin the telling.<sup>8</sup> Obviously, the speaker need not present one with an *independent* articulation of those reasons. They need not present the claims *q* and *r* ( $\neq p$ ) as reasons for accepting *p*. Rather, in telling one *p*, a speaker can *manifest* their possession of a piece of knowledge, where that piece of knowledge carries with it whatever reasons constitute it as knowledge. (As noted above, it might be allowed that a speaker can manifest a piece of knowledge that they do not possess, if their telling is appropriately dependent upon those who do possess it.) The genuine manifestation of knowledge in a telling, as understood here, requires that the telling be *constitutively* dependent on its epistemic bases: that very telling could not have occurred without its being based in the same way on knowledge. Insofar as it's possible for a speaker, in telling one that *p*, to manifest a piece of knowledge, it's thereby possible for them to manifest reasons sufficient for that knowledge. (Compare: insofar as it's possible for one to see water, it's possible for one to see Hydrogen and Oxygen, whether or not one can see Hydrogen or Oxygen independently of seeing water.) Hence, according to such

---

<sup>8</sup> McDowell 1980, 1981, 1994.

accounts, an appropriately based telling might make manifest reasons that tip the balance in favour of acceptance by deciding the issue.

However, even if we are prepared to endorse such an account, there are at least two apparent reasons for doubting its ability fully to explain how one can responsibly accept what one is told. First, doxastic responsibility requires that one's accepting what one is told way be based upon reasons. Those reasons must therefore be available to one in advance of one's accepting what one is told. And yet it's plausible that in cases in which reasons are made available to one by some source, one's initial access to those reasons is dependent upon one's faith in the source. It may be that once those reasons are accessible, they can then be used to consolidate one's initial grounds for faith in the source. However, they cannot serve as the initiating basis for faith. (This is not intended to constitute an objection to McDowell's *overall* account. I think that McDowell would accept that the aspect of his account characterized above requires supplementation. His overall account of doxastic responsibility incorporates additional elements, akin to those in the second account described below.)

Second, it is arguable that the claim that a speaker can manifest their knowledge to an audience by telling the audience something carries implausible implications concerning what a speaker ordinarily makes available to their audience. The claim requires that the reasons that are made available to an audience by one who speaks from knowledge are different from the reasons that are made available by one who doesn't. And if the reasons that are available in the good case are to enable an audience to meet the demands of doxastic responsibility, then those reasons must make a difference to how things are from the audience's perspective. Minimally, it seems to be required that an audience who is normally cognizant of a telling in a good case will thereby be put in a different epistemic position from an audience who is normally cognizant of a telling in a bad case. And it is natural to expect that requirement to have broadly phenomenological ramifications. That is, it is natural to predict that the way things seem to the audience in the good case will differ from the way things seem to the audience in the bad case. However, many theorists will be inclined to reject the claim that there is such a phenomenological difference between being in a good case and being in a bad case. Faulkner is amongst them:

...there is no significant phenomenological difference between being told something (by a speaker who knows what she tells) and being told a fiction. (Faulkner, 2011: 132)

Unless there is a way either to defend the phenomenological claim to which Faulkner is opposed, or to detach it from the proposed account, we will have to look elsewhere for an explanation of the doxastic responsibility of accepting what we are told.

A natural reaction to the first difficulty with the proposed account suggests a way forward. The difficulty was that, where access to reasons depends upon one's faith in a putative source for those reasons, those reasons are not able to ground one's initial faith in the source. But the difficulty is apt to generalize. If we assume that reasons for faith in a source must themselves have a source, it is natural to query the standing of faith in the latter source. To avoid the gloomy prospect of an interminable regress, we must therefore hold either that some sources may be trusted without reason, or that some reasons have no source. Crudely, we must hold that we are *pro tanto* entitled to faith in some putative sources. And having made that admission, the question naturally arises whether others' testimony is amongst the sources on which we are entitled to rely. If it is, then in the absence of reasons *not* to accept what someone tells us, we will be in a position responsibly to accept what they say.

Tyler Burge has developed an account on which we are *pro tanto* entitled to faith in others' testimony.<sup>9</sup> In outline, Burge's view is this. The presentation and uptake of testimony is based on the operations of a number of sources. For instance, a subject, *S* might come to know *p* on the basis of the operation of one source—their perceptual faculty—and then retain that knowledge on the basis of the operation of another source (or resource)—their memorial faculty. *S* might then make use of another source to manifest their knowledge in the presence of an audience *A*. That is, *S* might exercise their faculty of speech so as to tell *A* *p*—to present *p* to them as true. *A* is then in a position to grasp that which *S* told them, as having been presented by another as true, on the basis of the

---

<sup>9</sup> Burge 1993, 1997, 1998a, 1998b, 1999.

operation of *A*'s own speech faculty. If each of the sources that is involved in the transaction has functioned correctly, then *A* will be in the following position: if they responsibly accept what they grasp in that way—*p*, presented by another as true—they will thereby be accepting *p* on the basis of responsible and correct operations of various epistemic sources. Plausibly, that is a way for *A* to come to know *p* (at least on the assumption that *A*'s acceptance of what they are told is doxastically responsible). *A*'s acceptance of what they are told will be doxastically responsible just in case *A* is entitled to accept what they are told. *A* is *pro tanto* entitled to rely on the proper functioning of each of the various sources that participate in the transaction. So, as long as that *pro tanto* entitlement is not undermined, *A* can responsibly accept that which *S* tells them. *A* can come thereby to know *p*.

Faulkner objects to Burge's account in a way that mirrors, and bolsters, his phenomenological objection to McDowell's account. According to Faulkner, the basic difficulty for Burge's account resides in the presumed connection between belief (or knowledge) and its expression in speech. On Burge's account, we are entitled to rely upon the proper functioning of rational agents, and hence upon the proper functioning of their sub-systems. Suppose that there were a systemic connection between an agent's knowledge and its expression in speech, such that the function of the system was to make knowledge available through telling. In that case, an audience would be entitled to rely upon the proper functioning of the system, and so in a position responsibly to treat what they were told by a speaker as manifesting the speaker's belief. Put another way, audiences would be entitled to faith in speakers' sincerity.

However, as Faulkner emphasizes, speaking is a practical activity, undertaken by agents. The faculty of *practical* reason is therefore implicated in the transition from belief (or knowledge) to telling. Whether or not what a speaker says appropriately reflects what they believe depends, not only upon the proper operation of their faculty of speech, but also upon the dictates of the speaker's practical reasoning. On Faulkner's view, the question posed for practical reason is whether it would best serve the speaker's ends for them to speak their mind, or whether those ends would be better served by their speaking artfully—through lying, evasion, or some other form of insincerity. The answer to that question may

well be that the speaker's ends are best served by their speaking artfully. According to Faulkner, then, the connection between belief and speech is unsuitable to figure in Burge's account. Although an audience is entitled to rely on the proper functioning of each of the faculties implicated in a speaker's telling them something, that entitlement will ground faith in the speaker's sincerity only if the proper functioning of the faculties is inconsistent with artfulness. But the faculty of practical reason can function properly while exploiting the faculty of speech at the service of whatever the speaker's practical ends happen to be. Since those ends might mandate artful speech, the proper function of the system as a whole is consistent with artfulness. Hence, an entitlement to rely upon proper functioning cannot sponsor faith in a speaker's sincerity. Given suitable ends, a speaker might function perfectly and yet speak artfully.

Reflection on the role of practical reason in mediating the transition from belief to speech can serve to bolster Faulkner's phenomenological objection to McDowell's account. The objection, recall, was that there need be no accessible difference—no difference that impacts upon how things strike an audience—between a telling that is sincere and one that is insincere. Now one might counter that, although the difference need not everywhere be reflectively accessible to an audience, still an epistemically relevant difference can be made out. For one might argue that the appearance of a good case telling is constituted as a manifestation of knowledge, while the bad case telling presents what is merely a misleading appearance of being so constituted. However, that line requires that a good case telling amounts to the emergence of knowledge in action—i.e. that the telling could not have taken place had it not been a manifestation of the same piece of knowledge. And that arguably requires either that the expression of knowledge in speech be unmediated or that, if it is mediated, the mediator does not screen off the speech from its ultimate source in knowledge. However, as Faulkner characterizes the role of practical reason as mediating between belief and speech, it looks to sever the required connection between belief and speech. An episode of speech is shaped immediately by the operation of practical reason, where that operation may take any of a variety of configurations, depending upon the speaker's particular, idiosyncratic ends. Although some of those possible configurations will reflect the speaker's knowledge, and will lead to acts of speech that correspond with

that knowledge, not all will. So, at best, a good case telling is constituted as a manifestation of the immediate practical aims that it expresses. But those immediate practical aims are consistent with a variety of further ends and means. Hence, the telling is consistent with the speaker's having any of a variety of beliefs (or pieces of knowledge). Practical reason serves to screen off belief from speech. Hence, if Faulkner's objection to Burge were cogent, then the putative response to his phenomenological objection would be blocked.

## 5. Reason

Let's suppose for the moment that Faulkner's objections to Burge and McDowell are cogent. The supposition has two immediate consequences. First, and most obviously, we are owed an alternative account of how accepting what one is told in advance of positive grounds for doing so can meet the demands of doxastic responsibility. Faulkner's account of trust (a form of faith) as providing *evidence* of trustworthiness is designed to discharge that debt.<sup>10</sup> Second, we are forced to adopt a particular form of *externalism* with respect to knowledge that is acquired on the basis of accepting what one is told. I shall focus here on the second, less obvious consequence.

Consider a particular subject, *A*, who knows *p*, and consider all the reasons that figure in making it so that *A* knows *p*. Amongst those reasons, some may be reflectively accessible to *A*: they are such that, just by reflection, *A* is able to know that they have that reason. Let's call reasons that are reflectively accessible in that way *internalist* reasons. And let's call reasons that are not so accessible *externalist*. Now, Faulkner's objection to McDowell relies on three claims. First, it relies on the claim that any (normal) good case telling is paired with at least one matching bad case telling. (A minimal characterization of matching might proceed along the following lines: cases match just in case it is not reflectively accessible to one that one is not in a good case.) Second, it relies on the claim that no reflectively accessible reason is available in a good case that is not also available in every matching bad case. That is, it relies on the claim that the internalist reasons that are available in a good case cannot outstrip the reasons that are available in the worst matching

---

<sup>10</sup> Arguably, Faulkner's account is subject to one of the difficulties that Burge's account is designed to address. Suppose that one's trust in a speaker can serve as evidence for their trustworthiness, by conditioning them to be trustworthy. It's not clear how that evidence could ground the *initiation* of that trust.

bad case. The first two claims deliver the result that the internalist reasons available to *A* in a (normal) good case are consistent with *A* being in a bad case. That is, the internalist reasons that are available to *A* leave open that the telling with which they are presented is not knowledge-based. Hence, if *A* comes to know in the good case, that is in part because of the *externalist* reasons that are made available in the good case. It is the first two claims that impose a form of externalism on Faulkner's position. Although reasons can be transmitted from speaker to audience, internalist reasons cannot always be so transmitted. A speaker *S* may possess internalist reasons, but be blocked from transmitting those reasons to an audience *A*, because *A* would be unable reflectively to distinguish the case they're in from one in which *S* lacked those reasons.

The third claim on which Faulkner's objection to McDowell relies is that questions of doxastic responsibility must be settled by appeal only to internalist reasons. The first two claims aim to demonstrate the parity of internalist reasons across matching good and bad cases. The third aims to convert that result into an objection to the proposed account of doxastic responsibility.

There is a difficulty for Faulkner at just this point. It would be natural to expect facts about what one knows or doesn't know to play a role in determining whether or not one meets the demands of doxastic responsibility. Indeed, it would be natural to expect that to be one central reason for taking an interest in whether or not one knows something. However, with respect to the cases of knowledge that are of greatest interest to Faulkner—the normal cases in which one acquires knowledge from another—the expectation is not met. Some of the reasons that play an essential role in determining whether or not one knows are externalist reasons. And those externalist reasons are precluded from figuring in the determination of whether or not one's beliefs meet required standards of doxastic responsibility. On Faulkner's account, wherever one knows just on the basis of (responsibly) accepting what one is told, it is one's belief only, and not one's knowledge, that figures in determining whether one is, in further cases, being responsible in believing as one does.

It may be that Faulkner is willing to accept that result. If he is willing to accept it, then it would be good to hear more about what he takes the value, or explanatory function,

of knowledge to be, given that it is inefficacious with respect to questions of doxastic responsibility. Alternatively, however, he may wish to respond by amending one or another of the claims that figure in his objection to McDowell. First, he might consider dropping (or weakening) the internalist requirement on determinants of doxastic responsibility. That is, he might consider allowing that reasons that are not reflectively accessible to a subject may yet figure in determining whether or not they count as doxastically responsible in believing as they do. That would allow for a variant of the account considered above, on which one might be doxastically responsible in accepting what one is told because the telling manifested externalist reasons. However, one might well be unwilling so brutally to sever the connection between internalist reasons and the standards of doxastic responsibility. Second, then, Faulkner might consider dropping (or weakening) the parity principle, according to which the reasons that are reflectively accessible in a good case are restricted to those that are also accessible in any matching bad case. Taking that line would be to endorse McDowell's own way past the phenomenological objection.

If we're to pursue the second line, we'll need to reconsider Faulkner's objection to Burge. For in the absence of Burge's account, it would remain unclear how initial faith in a putative source of reasons can be doxastically responsible. And we would also lack a response to Faulkner's attempt to bolster his objection to McDowell by appeal to the practical nature of speech.

Faulkner's objection to Burge makes two large assumptions concerning practical reason. The first large assumption concerns the role of practical reason in mediating between knowledge and speech. The second concerns the nature of practical reason itself.

The first assumption is that, even in the good—or best possible—case, practical reason is responsible for the constitution of speech in a way that excludes the linguistic *manifestation* of knowledge. (Recall that, as understood here, the manifestation of knowledge in a telling requires that the telling be *constitutively*, and not merely causally, dependent upon the knowledge.) In outline, the assumed model is one on which practical reason and its sub-systems have access to what the subject knows, and may operate with due regard to that knowledge, but on which practical reason and its sub-systems shape the subject's activities in a way that is only indirectly responsive to that knowledge. In

particular, the immediate precursors of speech, and the speech itself, would have had the same nature even if it had not been based on knowledge. According to an alternative model, one form of activity available to a subject is the linguistic manifestation of knowledge. On the alternative model, there is a direct connection between knowledge and the power to act, so that action forged via that connection would be a manifestation of knowledge. One role of practical reason on the alternative model would be, not to *mediate* between knowledge and action, but rather to exploit the standing power of the system to manifest knowledge. On the alternative model, practical reason would be in a position to *trigger* manifestations of knowledge, and wouldn't be restricted to mirroring knowledge through its own repertoire of proprietary activities. If it were possible to develop and defend the alternative model, it would be possible at least to defend McDowell's position.<sup>11</sup> For in a good case, where practical reason triggers the manifestation of knowledge in a telling, what was triggered would differ in epistemic kind from any bad case telling.

Furthermore, with one addition, such an account might provide the resources to defend Burge's account. An addition is required because, as it stands, the alternative account leaves open that the system might function perfectly and yet fail to manifest knowledge. For the fact that the power to manifest knowledge is available to the system imposes no functional requirement on its exploiting that power. And we have as yet seen no reason for making it a requirement on the proper functioning of practical reason that it exploits any particular power. As things stand, the question whether practical reason should exploit that power, rather than exploiting different powers in order to issue artful speech, can be answered only by appeal to subjects' idiosyncratic ends. Hence, for all we have said to this point, the *pro tanto* entitlement to faith in proper functioning furnishes no grounds for faith in a speaker's sincerity. In order to shore up the account we would need to defend the claim that a function of the speech system is to manifest knowledge. It would follow that, in those cases in which practical reason intervenes between belief and speech in order to produce artful speech, the intervention would evince a malfunction in the speech system.

Faulkner's second assumption is that the nature of practical reason is such that its optimal functioning is consistent with its determining an agent to issue speech that is both

---

<sup>11</sup> McDowell 1981 attempts to articulate such a position.

insincere and apt to mislead. (Notice that it is not enough for Faulkner's purpose that practical reason should give rise to artful speech where that is made suitably manifest to the audience. In that case, there would be no obstacle to exploiting one's *pro tanto* entitlement to faith in sincerity in those cases in which a speaker's artfulness was not made manifest.) Faulkner doesn't explain why he holds that practical reason has such a nature, but his comments about the basic interests of a speaker and their audience are suggestive:

The testimonial situation, which is a conversation whose ostensible purpose is the giving and receiving of information, always involves a confluence of practical interests. ...[A]n audience's basic interest is learning the truth whereas a speaker's basic interest is being believed. That is, a speaker's basic interest, qua speaker, is not informing but exerting an influence. (Faulkner, 2011: 118–9)

Faulkner's idea appears to be that agents have, as their overriding practical interest, the furthering of their own egocentric ends. The audience has certain ends for the service of which they require certain information. The speaker has certain ends for the service of which they require their audience to take something to be true. The audience and the speaker therefore have different mediate ends. Moreover, the speaker's mediate end is shaped, not by the end of helping the audience, but rather by the end of exerting a cognitive influence on them. The speaker's end therefore may not coincide in its satisfaction conditions with the end of helping the audience.

The view of practical reason that Faulkner assumes may be correct. But it is not obviously so. For while it is certainly true that individual agents' ends may fit the pattern he describes, Faulkner's argumentative purposes require that they may do so without that evincing a malfunction in practical reason. And it is not obvious that a speaker's failure to respond to the epistemic needs of an audience would not constitute such a malfunction.

According to an alternative account, the basic function of practical reason is the selection and attainment of ends that are *good*. (We can leave open for present purposes whether or not the goodness of ends is taken to be determined independently of the nature of practical reason.) On the alternative account, practical reason would not merely be at the service of ends that were selected, independently of practical reason, by agents'

idiosyncratic needs and desires. Rather, one task for practical reason would be to select the right ends, those that are consistent with moral requirements. And a secondary task would then be to select means to those ends—mediate ends—in a way that was similarly consistent with those moral requirements. Where practical reason failed in its tasks—where it selected ends or means that were inconsistent with the demands of morality—that would constitute a malfunction.

Suppose that such an account is correct. There is arguably a *pro tanto* moral requirement to the effect that one should do one's best to service an audience's epistemic ends; or, at least, that one should do one's best not to disrupt their capacity to attain such ends.<sup>12</sup> (Such a *pro tanto* requirement might be undermined if, for example, an audience's epistemic ends were themselves inconsistent with moral requirements. Consider, for example, Kant's example of a person with known murderous intent who demands that one tell them whether or not a potential victim is at home.<sup>13</sup>) One way of failing to do one's best not to disrupt an audience's epistemic ends would be to mislead them into taking to be true something that one didn't oneself take to be true. And one way of doing that would be to tell the audience something that was not appropriately based on what one believes—that is, to speak artfully or insincerely—without making manifest what one were up to. It is therefore a plausible *pro tanto* moral requirement that one should avoid telling an audience something where the telling is misleading as to its artfulness or insincerity. Putting that together with the proposed account of practical reason would then deliver the result (as plausible) that misleadingly insincere or artful telling would evince a malfunction in practical reason.<sup>14</sup>

It would of course take extensive reflection and argument properly to assess the prospects of such an account. And it is far from obvious what the fruits of such an assessment would be. However, if such an account could be defended, it would feed into an account on which there is a *pro tanto* entitlement to accept what one is told. For (we are assuming) there exists such an entitlement to faith in the proper operation of subject's

---

<sup>12</sup> Korgaard 1986 is a useful discussion of Kant's defence of such a moral requirement.

<sup>13</sup> See Kant 1797.

<sup>14</sup> At this level of generality, the resulting position has a recognizably Kantian caste. For relevant discussion of Kant's views in this area, see Gelfert 2006, Shieber, 2010.

faculties and sub-faculties. That entitlement to faith in proper operation plausibly extends to the faculty of practical reason. And what we have just sketched is an account on which a telling would manifest the proper operation of practical reason only if it were not misleadingly artful or insincere. So, where a telling appears—to reasonably demanding standards—to be sincere, we are *pro tanto* entitled to take it to be sincere. We therefore have a second potential plug for the hole in Burge’s account revealed by Faulkner’s argument from the practical nature of speech. We would be entitled to faith, not only in others’ epistemic capacities, but also in the goodness of their wills.<sup>15</sup>

## 6. Conclusion

My aim here has been to outline some questions that remain open concerning the terrain partly mapped in Faulkner’s provocative book. I’ve pursued that aim in two ways. First, I’ve tried to indicate some areas that remain to be explored. The two main questions here are these. (1) What is the nature of the distinction between faith and knowledge? (2) What is the role of faith in the justificational architecture of testimonial belief? Second, I’ve tried to indicate some ways in which some of the paths that Faulkner argues to be impassable might yet be navigated. The four main questions here are these. (3) Should we accept externalism about testimonial knowledge? (4) What is the function of knowledge, as opposed to other forms of cognition? In particular, what is its function with respect to conditions on doxastic or practical responsibility? (5) What is the role of practical reason in the constitution of belief based speech? (6) Is misleadingly artful telling consistent with the proper function of practical reason?

**Guy Longworth**

*University of Warwick*

G.H.Longworth@warwick.ac.uk

---

<sup>15</sup> Cp. Ross 1986.

## References

- Burge, T. (1993) 'Content Preservation', *Philosophical Review*, 102: 457–488.
- (1997) 'Interlocution, Perception, and Memory', *Philosophical Studies*, 86: 21–47.
- (1998a) 'Computer Proof, Apriori Knowledge, and Other Minds', *Philosophical Perspectives*, 12: 1–37.
- (1998b) 'Reason and the First Person', in B. C. Smith, C. MacDonald, and C. Wright (eds.) *Knowing Our Own Minds: Essays on Self-Knowledge*, Oxford: Clarendon Press.
- (1999) 'Comprehension and Interpretation', in Lewis E. Hahn (ed.) *The Philosophy of Donald Davidson*, Chicago and La Salle, Illinois: Open Court.
- Chigwell, A. (2007) 'Belief in Kant', *Philosophical Review*, 116 (3): 323–360.
- Dretske, F. (1970) 'Epistemic Operators', *Journal of Philosophy* 67: 1007–1023.
- Faulkner, P. (2011) *Knowledge on Trust*. Oxford: Oxford University Press.
- Gelfert, A. (2006) 'Kant on Testimony', *British Journal for the History of Philosophy*, 14(4): 627–52.
- Graham, P. (2000) 'Conveying Information', *Synthese*, 123 (3): 365–92.
- Kant, I. (1781/1787 [1998]) *Critique of Pure Reason*. Translated and edited by P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.
- (1785 [1996]) "What Does it Mean to Orient Oneself in Thinking?" in *Religion and Rational Theology*. Translated and edited by A. W. Wood and G. di Giovanni. Cambridge: Cambridge University Press.
- (1797) 'On a Supposed Right to Lie from Philanthropy', in *Kant: Practical Philosophy*. Translated and edited by M. J. Gregor. Cambridge: Cambridge University Press.
- Korsgaard, C. (1986) 'The Right to Lie: Kant on Dealing with Evil', *Philosophy and Public Affairs*, 15(4): 325–49. Reprinted in Korsgaard 1996.
- (1996) *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

- McDowell, J. (1980) 'Meaning, Communication, and Knowledge', in Z. van Straaten ed. *Philosophical Subjects: Essays Presented to P. F. Strawson*, Oxford: Clarendon Press. Reprinted in McDowell 1998.
- (1981) 'Anti-Realism and the Epistemology of Understanding', in H. Parret and J. Bouveresse eds. *Meaning and Understanding*, Berlin: Walter de Gruyter. Reprinted in McDowell 1998.
- 1994. 'Knowledge By Hearsay' in B. K. Matilal and A. Chakrabarti eds. 1994 *Knowing from Words: Western and Indian Philosophical Analyses of Understanding and Testimony*. Kluwer: Dordrecht: 195–224. Reprinted in McDowell, 1998.
- 1998. *Meaning, Knowledge, and Reality*. Cambridge, Mass.: Harvard University Press.
- Nozick, R. (1981) *Philosophical Explanations*. Oxford: Clarendon Press.
- Ross, A. (1986) 'Why do we believe what we are told?', *Ratio* 27: 69–88.
- Shieber, J. (2010) 'Between Autonomy and Authority: Kant on the Epistemic Status of Testimony', *Philosophy and Phenomenological Research*, 80(2): 327–48.
- Sherman, B. and Harman, G. (2011) 'Knowledge and assumptions', *Philosophical Studies*, 156: 131–40.
- Stevenson, L. (2003) 'Opinion, Belief or Faith, and Knowledge', *Kantian Review*, 7: 72–101.
- Stine, G. (1976) 'Skepticism, Relevant Alternatives and Deductive Closure', *Philosophical Studies* 29: 249–261.
- Vogel, J. (1990) 'Are there counter-examples to the closure principle?' in M. D. Roth and G. Ross eds. *Doubting: Contemporary Perspectives on Skepticism*, Dordrecht: Kluwer.
- Williamson, T. (2000) *Knowledge and its Limits*. Oxford: Oxford University Press.

## KNOWLEDGE ON AFFECTIVE TRUST

**Arnon Keren**

Epistemologists of testimony widely agree on the fact that our reliance on other people's testimony is extensive. However, they have not always paid enough attention to the fact that our reliance on testimony is not only extensive, but also varied. That is, there is more than one way in which we can rely on a speaker's testimony to form a belief: Sometimes we treat a person's testimony that  $p$  as we treat any other piece of evidence. At other times, we just trust the speaker for the truth of what she says. Arguably, the latter is the normal way of forming a belief on the basis of other people's testimony. And while Paul Faulkner's important book presents an account of how we can obtain knowledge in both these ways, its focus, as its title suggests, is on knowledge obtained through trust.

Faulkner is one of a number of epistemologists who have attempted in recent years to explain how we can obtain knowledge by trusting others (Faulkner 2007; Fricker 2006; Hinchman 2005; Keren 2007; McMyler 2011; Moran 2005). It is hard to deny that knowledge can be obtained in this way, and Faulkner's attempt to explain how this can be possible is both original and sophisticated. It thus constitutes a significant contribution both to the epistemology of testimony, and to the philosophical study of trust.

Faulkner's book is divided into three major parts. The first (chapter 1) presents a problem—the problem of cooperation—and draws a conclusion from the analysis of the problem that underlies much of the discussion throughout the book: This is the claim that it would be reasonable for an audience to trust a speaker only if she has some positive reason to think that the speaker is trustworthy or cooperative.

This claim is at the basis of some of Faulkner's central arguments in the second part of the book (chapters 2-5) in which he discusses and rejects reductive and non-reductive approaches to testimony, and defends instead a distinct theoretical view in the epistemology of testimony: one that combines some of the more plausible elements of both, while rejecting the less plausible ones. Thus, with the non-reductionist he claims that testimonial

knowledge is transmitted knowledge. Accordingly, a necessary condition for obtaining *testimonial* knowledge by trusting a speaker is that the speaker herself knows. And it is in virtue of its being transmitted knowledge that testimonial knowledge is a distinctive kind of knowledge. At the same time he agrees with the reductionist that one can obtain knowledge by trusting a speaker only if one has reasons, epistemic reasons, for trusting her. Thus he rejects the non-reductionist claim that we have a default entitlement to accept the testimony of a speaker.

The third part of the book (chapters 6 and 7) proposes a positive solution to the problem of cooperation, explaining how we can have reasons to trust a speaker. Faulkner adopts a non-doxastic account of a form of trust—*affective trust*—according to which, to trust a person, one need not believe that the person is trustworthy. Moreover, according to Faulkner's account, to be rational in trusting a person, one need not have evidence that the trusted person is trustworthy. While Faulkner's account of trust shares important features with other accounts of trust found in the literature, e.g., in Holton (1994), Faulkner's account is much more developed, and his explanation of how trust can be rational, and how beliefs formed through trust can be epistemically warranted, is highly original. Ultimately, he claims, that affective trust is rationally self supporting, and that it is this trust itself which provides an audience with an epistemic reason for her belief in the testimony of the speaker.

There is a lot that is going on in this rich book. In my discussion, I will therefore only address some key aspects of his study of knowledge formed through trust, focusing on those aspects with which I disagree. I will first raise some worries about Faulkner's analysis of the problem of cooperation. Ignoring much of Faulkner's discussion of reductive and non-reductive position, with which I largely agree, I will then turn to a discussion of Faulkner's suggested solution of the problem. As I will argue, this latter part of the book raises a number of fundamental worries. It is therefore doubtful whether Faulkner has presented us with an adequate explanation of how knowledge is obtained through trust.

### **The Problem of Cooperation**

Faulkner opens his book by presenting what he calls "the problem of cooperation". From his analysis of the problem, he concludes that it would be reasonable for an audience *A* to trust a speaker *S*, only if *A* has some positive reason to think that *S* is trustworthy or cooperative. This claim is at the basis of Faulkner's arguments against both reductive and non-reductive approaches to testimony. And it lies at the basis of the problem that Faulkner's trust theory of testimony is designed to solve: that of explaining the nature of the positive reasons that make it reasonable for *A* to accept *S*'s testimony, and in particular, to trust *S* for the truth of what she says.

Faulkner presents the problem of cooperation by drawing a parallel between two choice situations in contexts where one subject can decide whether to trust another. The first is a purely practical decision, of the kind subjects need to make in the 'Trust Game'. In this game, an 'investor' is given an initial sum of money (say £10), and is asked to choose whether to cooperate, and to transfer some or all of it to the 'trustee', or whether to defect and keep it all to himself. Whatever sum is transferred is then multiplied, say by a factor of 4, and the trustee is asked to decide whether to cooperate, and give back all or part of the multiplied sum to the investor, or to defect and keep it all to herself. Given very minimal assumptions about the two parties' preferences, it is plausible to presume, first, that the trustee would prefer to keep whatever money is transferred to her; and, second, that the best outcome for the trustee—where the investor cooperates and the trustee defects—is also the worst outcome for the investor. Because of this, Faulkner concludes, it would be unreasonable for the investor to cooperate, and to make the initial transfer, unless she has reason to think that the trustee would cooperate.

Faulkner then draws a parallel between the trust game and the 'Testimony Game', to argue that a similar conclusion applies to the case of testimonial encounters: audience *A* is unreasonable in accepting speaker *S*'s testimony unless *A* has positive reasons for thinking that *S* is cooperative or trustworthy. Faulkner bases this conclusion on a *parity claim*, according to which, given some very minimal assumption about the preferences of the parties, the testimony game has the same payoff structure as the trust game: the best option for the potentially trusted party, the speaker, is the worst outcome for the potentially

trusting party, the audience (2011, 6). He then goes on to claim that this parallel implies that, as in the trust game, "one can equally draw the conclusion that [in the testimony game] it is not reasonable to trust without a supporting reason that rationalizes trust" (2011, 6).

While I ultimately agree with Faulkner's conclusion, I have worries about his parity claim, and therefore with his argument for this conclusion. Why should we think that in the testimony game, as in the trust game, the best outcome for the speaker is the worst for the audience? The good that the audience hopes to obtain in the testimony game—information—is very different from that which the investor hopes to obtain in the trust game. Unlike money, information is a good whose consumption is non-rivalrous, so that providing the audience with that good does not generally reduce the benefit available to the speaker from its use. So there is no reason to presume that the speaker should generally prefer to provide the audience with misinformation rather than with information. Of course, there are testimonial encounters, such as that between a car dealer and a potential buyer, in which the speaker might have an interest in misleading the audience. However, Faulkner's claim is that such a conflict between the interests of speakers and audiences is not peculiar to some particular testimonial interactions. Instead it is a general feature of testimonial encounters, as exhibited by the testimony game.

While Faulkner's parity claim does not seem to be true if the audience is represented as merely seeking to form accurate, true beliefs, it might be thought that the claim holds if we conceive of her instead as seeking knowledge. Indeed this appears to be how Faulkner attempts to base his parity thesis: The ranking of outcomes attributed to the audience is supposed to be true "to the extent that an audience's interest is epistemic". Our interests as audience is in "learning the truth," which, I take it, involves more than just obtaining a true belief (2011, 5); it involves obtaining knowledge. Arguably, to obtain knowledge requires a speaker who does not only tell the truth when it suits him. It requires a speaker who is trustworthy. Accordingly, Faulkner claims that the best outcome for the audience is obtained when she trusts a trustworthy speaker. But, he claims, the best option for the speaker is to be believed without being trustworthy, and this is the worst outcome for the audience.

However, the assumption that the goal of the audience is the attainment of knowledge only makes matters worse for Faulkner's parity claim. For it points at a fundamental reason for thinking that the trust game and the testimony game cannot have a similar payoff structure. For if what the audience is after in the testimony game is knowledge, then her payoff depends not only on the speaker's decision and on her own; it also depends on whether her trust of the speaker is reasonable. Arguably, even if the speaker happens to be trustworthy, and the audience trusts her, the audience will not end up with knowledge if it was not rational for her to trust the speaker. In contrast, in the trust game, the outcomes for the parties depend only on their decisions, and not on their reasons for deciding as they did.

Whether or not Faulkner is correct in claiming that it is unreasonable to trust a speaker without positive reasons supporting such trust, the claim is not implied by his conclusion about the trust game, because of the disparity between the games: In the trust game, payoffs are not a function of the parties' reasons for their choices, but in the testimony game they are. As a result, the argument from cooperation fails, and some of Faulkner's key arguments against reductive and non-reductive approaches to testimony are lacking. This is so in as much as these arguments attempt to show that reductive and non-reductive approaches to testimony fail to adequately address the problem of cooperation. Thus, a central objection to the reductive theory is that the reductive solution to the problem of cooperation is too restrictive, in that it misses the central reason for trusting the testimony of the speaker: that we trust the speaker (2011, 53-55). Similarly, the central objection made against non-reductive approaches to testimony, such as those offered by Burge and McDowell, is that the acceptance principles offered by them are inconsistent with the principle of reasonable uptake (R), which is supposedly supported by the argument from cooperation: the principle according to which "an audience *A* is warranted in believing [a speaker's testimony] that *p* if and only if *A*'s other attitudes make it reasonable for *A* to believe that *p*" (Faulkner 2011, 119). Since the argument from cooperation fails to establish (R), proponents of the non-reductive view will probably not be convinced by this argument.

### **The Trust Theory: Solving the Problem of Cooperation**

While I have some reservations about the way Faulkner relies on the argument from cooperation to reject reductive and non-reductive approaches to testimony, I think that he is ultimately correct both in his rejection of these approaches, and in his reasons for rejecting them. We can obtain knowledge by trusting speakers, and both approaches fail in explaining how we do so: the reductive theory is unable to address what is distinctive about knowledge based upon trust; and the non-reductive theory is mistaken in suggesting that we do not need positive reasons to trust. So Faulkner is correct in arguing that a main task of the epistemology of testimony, perhaps *the* main task, is that of explaining how such trust can be reasonable. Whether or not his attempt to meet this task is successful, Faulkner's book makes an important contribution to the epistemology of testimony by first focusing our attention on this challenge; and second, by presenting an original approach in the epistemology of testimony, at the heart of which is a sophisticated theory which attempts to meet this challenge. Whether or not Faulkner's trust theory succeeds in doing so, his discussion of the theory clearly advances our understanding of the theoretical options available to us, and of the challenges that we must face on our way towards a solution of the problem.

There are a number of elements in Faulkner's solution to the problem. At the heart of the solution is an account of a thick form of trust, *affective trust*, which Faulkner distinguishes from a thinner notion of *predictive trust*. The main difference between the two forms of trust is that they involve two very different notions of expectation. In trusting *B* to  $\Phi$  in the predictive sense, we expect that *B* will  $\Phi$ . In trusting *B* to  $\Phi$  in the affective sense, our expectation of *B* that she  $\Phi$ 's is a normative expectation. We expect this of her. This kind of normative expectation is to be cashed out in terms of the reactive attitudes we shall be prepared to feel if *B* fails to  $\Phi$ . The main task Faulkner sets to himself is that of explaining how we can obtain knowledge by affectively trusting speakers.

The first important element of the suggested solution to the problem of cooperation is a claim about the internalization of norms of trust. Here Faulkner's suggested solution builds on one proposed by Bernard Williams (2002), but also departs from it in a significant way. Williams suggests that as speakers we have internalized the value of sincerity, and

that it is our intrinsic valuation of sincerity that explains the reliability of testimony. His suggestion is that audiences are then rational in trusting speakers to the extent that they have grounds for believing that speakers would be trustworthy. Faulkner rejects the latter element of this solution, arguing that Williams' account of what makes it reasonable for audiences to believe speakers' testimony is at odds with the claim that we are trustworthy as speakers because we intrinsically value sincerity (Faulkner 2011, 177). Instead, he suggests, as audiences we trust speakers because we have internalized norms of trust requiring of us that we trust speakers, in the affective sense, just as we are trustworthy as speakers because we have internalized norms of cooperativeness in conversation.

It is not entirely clear however what kind of role the internalization of norms of trust is supposed to have within the solution of the problem of cooperation. As said above, Faulkner objects to Williams's reductive account of what makes it reasonable for audiences to believe speakers' testimony. He therefore replaces a key element in Williams account—the claim that audiences are rational in trusting speakers when they have grounds for believing that speakers are trustworthy—with the claim that audiences internalize norms of trust. But it is hard to see how the latter claim can play the kind of role played by the rejected claim: that of explaining the rationality of beliefs formed by trusting speakers. The problem emerges from a second fundamental difference between the trust game and the testimony game: This difference lies in the relations between the rationality of choices and the valuation of outcomes. When it comes to the decision how to act, the practical rationality of one's action can be a function of one's own valuation of possible outcomes. In contrast, the epistemic rationality of one's belief, and hence its epistemic status, is arguably not a function of one's own valuation of knowledge, or of various ways of forming beliefs. Thus, in the trust game, if the trustee cares more about the investor's well-being than about her own, or if she intrinsically values performing acts of kindness, then it might be perfectly rational for her to give back all the money transferred to her by the investor. If, on the other hand, she does not place any extraordinary value on the performance of acts of kindness, then giving all the money back to the trustee would arguably not be reasonable for her. In contrast, the audience's valuation of knowledge and trust does not seem to matter for the epistemic rationality of her trusting the speaker. That a thinker does not value

knowledge and accuracy as we do does not mean that she can be *epistemically* rational in forming beliefs contrary to the evidence. Similarly, the fact that a thinker intrinsically values the formation of beliefs in certain ways—by consulting the stars, or by trusting a speaker—does not seem to matter at all for the epistemic rationality of beliefs formed in these ways.

It may therefore seem unclear what role audiences' internalization of norms of trust is supposed to play within an explanation of the rationality of trust and of beliefs thus formed. Speakers' internalization of norms of sincerity can make it rational for them to speak truthfully, even in situations where speaking truthfully would not have been rational otherwise. However, unlike audiences' having evidence about speakers' internalization of norms of trustworthiness, it is difficult to see how audiences' internalization of norms of trust can make a difference to whether it is *epistemically* rational for them to trust speakers. Having rejected the idea that it is evidence that makes our trust rational, Faulkner's alternative suggestion may therefore appear to be irrelevant to the rationality of audiences' trust.

Perhaps Faulkner's idea is this: an explanation of knowledge obtained through (affective) trust should do two things. It must first explain why, faced with the problem of cooperation, audiences do indeed form beliefs by trusting speakers, in the affective sense. Call this the causal-motivational part of the explanation. The second part of the explanation is not causal, but normative, and addresses the epistemic status of beliefs thus formed: At this stage what needs to be explained is how beliefs thus formed can have the epistemic status of knowledge. It might be suggested that the discussion of the internalization of norms of trust plays a role only in the causal-motivational part of the explanation. The normative part of the explanation is done elsewhere (in chapter 6), where Faulkner explicitly sets out to explain how affective trust can be rationally-self supporting, and how affective trust can be epistemically rational.

This appears to be a plausible interpretation of Faulkner's suggested solution. There are, however, reasons for doubt about both parts of the explanation. The first problem involves the causal-motivational part. Our internalization of norms of trust may explain why we trust speakers, in the affective sense. But this is only part of an explanation of why

audiences form beliefs by trusting speakers. What must also be explained is why audiences who affectively trust speakers also believe what speakers tell them. But this part of the explanation is missing from Faulkner's account. And it is not clear whether this is something that Faulkner's account can adequately explain. This is because affective trust, unlike predictive trust, does not seem to involve belief in the trustworthiness of the trustee. Predictive trust involves the belief that the trusted party would do what she is trusted to do, and so it is clear why an audience who predictively trusts a speaker to speak knowledgeably and honestly would believe the speaker's testimony (provided that she understands what the speaker tells her). However, affective trust, on Faulkner's account, does not seem to involve the *belief* that the trusted party would do what she is trusted to do. Instead it involves the *presumption* that the trusted party would do so. But when audiences merely presume that speakers are trustworthy without believing this, why should we expect this presumption to result in their *believing* that the speaker's testimony is true, and not merely in their *presuming* that the testimony is true? Why would trusting the speaker invariably involve believing what she says (provided adequate understanding)? After all, we often presume that a defendant is innocent without forming the belief that she is. Faulkner needs to explain not only why trusting a speaker results in belief, but why it invariably has this result. And he must do so while maintaining the distinction between predictive and affective trust. It is not clear how this can be done.

In previous writings (such as Faulkner 2007), Faulkner in fact denied that affectively trusting a speaker always involves believing her testimony. Instead, he suggested there, it involves accepting the speaker's testimony as true, where acceptance, as suggested by Cohen (1992), is to be distinguished from belief, *inter alia*, in being under our direct voluntary control. This allowed Faulkner to claim that we have a kind of voluntary control over our trust, that we don't have over what we believe (2007, 894). But in *Knowledge on Trust* Faulkner withdraws from the claim that trusting a speaker is consistent with not believing what she says—and rightly so. For in as much as our reactive attitudes are a relevant indication, they seem to suggest that what speaker *S* expects of audience *A*, when she invites *A* to trust her, is not merely that *A* accept her testimony as true. "In telling *A* that *p*," he writes, "*S* will expect to be believed." (2011, 182). Accordingly, "we are liable to

resent...audiences who do not believe us...and [as audiences, to] feel pressure to believe what speakers tell us" (180). Faulkner is therefore right to claim that "When acceptance is motivated by an attitude of trust—when it is a case of trusting—it issues in belief. The act of trusting testimony is the uptake of testimony" (23). What is left unclear is how his account of affective trust can explain this invariable relation between trusting a speaker, and believing her testimony.

Perhaps Faulkner can meet this challenge by suggesting that affective trust involves having *both* normative expectations of the trusted person, and, in addition, the belief that the trusted person will be trustworthy. Indeed some passages in Faulkner (2011) seem to be consistent with such an interpretation. Thus Faulkner writes of the expectation involved in affective trust that "this expectation is *more than* a statement of our subjective probabilities" (146; emphasis mine). And if affective trust involves believing that the trusted party is trustworthy, then it is no mystery why affectively trusting a speaker should result in believing that her testimony is true. However, there are ample indications that Faulkner does not conceive of affective trust as involving belief in the trustworthiness of the trusted party. Faulkner does not conceive of the presumption involved in trust as a kind of belief: It is a different mental state, which like acceptance, is under our voluntary control in ways that belief is not (149-150). It is therefore unclear how Faulkner's account of affective trust can adequately explain the systematic relation between trusting a speaker and believing what she says.

### **The Epistemic Rationality of Trust**

Let us suppose that in spite of what was said in the previous section, Faulkner's account of affective trust can be part of an adequate explanation of the fact that we form beliefs on the basis of trust. We must still ask whether the account of affective trust can serve in the way suggested by Faulkner within an explanation of the epistemic status of belief formed in this way. Can we explain why beliefs formed by affectively trusting a speaker often count as knowledge?

Faulkner suggests a highly original explanation of the epistemic status of trust-based belief. It is original, first, in that it departs from most reductive and non-reductive

accounts, combining plausible elements of both, while rejecting less plausible ones: claiming with the reductionist that *A* can obtain knowledge by trusting *S* only if *A* has epistemic reasons for trusting *S*, and with the non-reductionist that testimony serves to transmit knowledge. It seems to me that Faulkner is correct in both of these claims.

However, it is Faulkner's account of the epistemic rationality of trust which is, in mind, the most original part of the book. There are two parts to Faulkner's account. First, he argues that it is *A*'s trust of *S* that provides *A* with an epistemic reason for believing *S*'s testimony that *p*, in the minimal sense that it makes *p* subjectively probable for *A* (153-4). Second, he argues, when the speaker is in fact trustworthy, the fact that the audience trusts the speaker is potential evidence that *p*, and thus epistemically warrants the audience's belief that *p* (154-159).

I have worries about the plausibility of both parts of this explanation. First, I have doubts about Faulkner's suggestion that trust provides *A* with an epistemic reason to believe *S*'s testimony because it makes it subjectively probable that the testimony is true. Faulkner's idea is that in affectively trusting *S*, *A* accepts, or presumes, that *S* will be motivated to tell the truth because *S* recognizes *A*'s dependence on *S*. And accepting this makes it probable for *A* that *S*'s testimony that *p* is true. The problem with this is that Faulkner does not explain how we are to understand this notion of presumption, and how presumptions can justify beliefs. If we are to conceive of a presumption as a kind of belief backed by a presumptive epistemic right, then it would perhaps be clear how such a presumption can justify a belief. But were this Faulkner's position then his position would not seem to differ from that of the non-reductionist. If, on the other, the mental state ascribed to an audience who presumes that the speaker is trustworthy is similar to that of a judge who presumes that the defendant is innocent because the law requires this of him, then it is not at all clear how such a presumption can make any belief *epistemically* reasonable, not even in a minimal sense.

In any case, as Faulkner is well aware, even if trusting the speaker would have made *A*'s belief in the truth of *S*'s testimony epistemically reasonable in the above minimal sense, this might not suffice to render it epistemically warranted. Faulkner argues, however, that when *S* is in fact trustworthy, then the fact that *A* trusts *S* epistemically warrants *A*'s belief

in *S*'s testimony. For in such a case, he argues, the fact that *A* trusts *S* is potential evidence for the truth of *S*'s testimony. And if so, then the fact that *A* trusts *S* provides *A* with potential evidence that the testimony is true, and hence renders *A*'s belief epistemically warranted. Faulkner argues for the first claim based on an account of potential evidence developed by Achinstein (1978). Essentially, the claim is that *A*'s trust is potential evidence for the truth of *p*, if two conditions hold—both of which are satisfied when *S* is trustworthy: If the fact that *A* trusts *S* makes it sufficiently objectively probable that *p* is true; and if it is sufficiently objectively probable that there is an explanatory relation between *A*'s trust and the truth of *p*.

However, the question is whether the fact that her trust is potential evidence for *p* suffices to warrant *A*'s belief that *p*. The fact that it is potential evidence for *p* would surely warrant *A*'s belief that *p* were this belief well-based on this evidence. However, even if *A* believes that *p* because she trusts *S*, and her trusting *S* is potential evidence for *p*, it is not at all clear that it would be correct to say of her belief that it is based on this evidence, let alone well based. Suppose, for the sake of illustration, that the following is true of a certain type of cancer patients: if patients of this kind have a desire to survive their illness, then this desire not only causes them to believe that they will survive it; it also sufficiently increases their chances of survival, and would play a role in explaining their survival should they survive, so that having the desire is potential evidence for the proposition that they will survive. Arguably, that such a patient desires to survive does not provide her with an epistemic reason for believing that she will survive, and does not suffice to warrant such a belief. While different epistemologists would make different suggestions about what else is required to make such belief warranted, most would agree that the following two facts do not suffice to make *A*'s belief that *p* warranted: That *A*'s having mental state *M* is potential evidence that *p* and that *A*'s having *M* is the cause of *A*'s belief that *p*. More than that is required for epistemic warrant. Some might suggest that *A* must also be justified in believing that *M* is evidence for *p*. Others, that *A* must in some sense be reliably sensitive to *M*'s being potential evidence for *p*, or causally related to *p*. The worry is that both types of additional conditions might not be satisfied in the case where an audience affectively trusts a trustworthy speaker.

It seems to me that this is a genuine worry in the case of affective trust. If the audience trusts the speaker because she has internalized norms of trust (regardless of whether her trust is evidence for the speaker's trustworthiness), and if trust invariably leads to belief, then even in cases where the audience's trust happens to be evidence for the truth of the belief, that it is evidence is not the audience's reason for belief. It is not the case that the audience believes that  $p$  because she has this evidence, or because she believes she has evidence; and she would believe the same regardless of whether her trust is evidence for her belief. It therefore seems that more needs to be said in order to argue that the audience's trust epistemically warrants her belief.

Moreover, it is not clear that this worry can be set aside by pointing to a parallel with perceptual appearances. True, appearances might be misleading. And it is an epistemological commonplace that thinkers cannot always distinguish between misleading perceptual appearances and veridical ones; and that the fact that a thinker would form a perceptual belief regardless of whether a perceptual appearance is potential evidence for the truth of the belief does not entail that perceptual appearances do not provide epistemological reasons for belief. However, the case of affective trust seems to be significantly different. Thinkers do not invariably form the belief corresponding to their perceptual appearances; trusting a speaker, in contrast, invariably involves believing the speaker's testimony. So the sense in which affective trust seems to involve forming a belief regardless of whether such trust constitutes evidence for the belief is not the sense in which perceptual appearances may lead us to form a belief regardless of whether they constitute evidence for it.

It therefore seems to me that Faulkner's arguments against alternatives to his trust theory are more convincing than his positive account of how knowledge is obtained through trust. Faulkner is correct in arguing that reductive and non-reductive approaches in the epistemology of testimony fail to adequately explain how knowledge is obtained through trust. However, I doubt if his account of affective trust provides us with a successful explanation of this. It is doubtful whether his account explains why we believe the testimony of speakers when we affectively trust them; and it is doubtful whether it explains why beliefs thus formed often constitute knowledge. Further discussion is required

to determine whether the gaps in the explanation can be filled. However, given the worries raised here, I doubt if any non-doxastic account of trust which like the account of affective trust, suggests that trusting a person need not involve belief in her trustworthiness, can successfully meet the challenge.

**Arnon Keren**

*University of Haifa*

akeren@research.haifa.ac.il

## References

- Achinstein, P. (1978) 'Concepts of Evidence'. In Achinstein, P. (ed.) *The Concept of Evidence*. Oxford: Oxford University Press.
- Cohen, L. J. (1992) *An Essay on Acceptance and Belief*. Oxford: Clarendon Press.
- Faulkner, P. (2007) 'Telling and Trusting', *Mind*, 116: 875-902.
- Faulkner, P. (2011) *Knowledge on Trust*. Oxford: Oxford University Press.
- Fricker, E. (2006) 'Second Hand Knowledge', *Philosophy and Phenomenological Research*, 73: 592-618.
- Hinchman, E. S. (2005) 'Telling as Inviting to Trust', *Philosophy and Phenomenological Research* 70: 562-587.
- Holton, R. (1994) 'Deciding to Trust, Coming to Believe', *Australasian Journal of Philosophy*, 72: 63-76.
- Keren, A. (2007) 'Epistemic Authority, Testimony and the Transmission of Knowledge', *Episteme*, 4: 368-381.
- McMyler, B. (2011). *Testimony, Trust, and Authority*. Oxford: Oxford University Press.
- Moran, R. (2005) 'Getting Told and Being Believed', *Philosophers' Imprints*, 5: 1-29.
- Williams, B. (2002) *Truth and Truthfulness*. Princeton: Princeton University Press.

## CAN TRUST ITSELF GROUND A REASON TO BELIEVE THE TRUSTED?

Edward S. Hinchman

Trust in testimony is curiously self-reflexive. As Paul Faulkner emphasizes in *Knowledge on Trust*, when we trust a testifier, we rely not simply on her, as we might if we merely overheard her, but specifically on her attitude toward the fact that we are relying on her. This distinctive form of reliance explains the second-personal nature of the testimonial relationship: when you trust someone's testimony, you rely not merely on her attitude toward the proposition that she asserts but on her second-personally directed attitude toward you yourself as the one whom she is addressing. It creates space for the reasonability of trusting a speaker who lacks a track record of reliability, since you may be in position to gauge how her present conscientiousness and responsiveness to your epistemic needs outweighs any history of error or insincerity. But it makes explaining testimonial warrant trickier than explaining the warrant that we may derive from observing forms of behavior in which we are not so intimately implicated.

Thus far, I completely agree with Faulkner. Though the claim needs a fuller vindication than I'll attempt to provide, I'll take for granted that you can acquire an epistemic reason to believe that *p* by accepting a speaker *S*'s testimony that *p* even when *S* has a track record of error or insincerity. How might you acquire such a reason? Consider how a different reason might arise on *S*'s side of the transaction. Just as you can make a clock reliable by fixing its gears, so you can make a speaker reliable by giving her a reason to tell you the truth.<sup>1</sup> How can you give someone a reason to tell you the truth? I agree

---

<sup>1</sup> It would amount to an odd view of testimonial trustworthiness to insist that *S* could not become trustworthy until she had established an appropriate history. Though we cannot inspect the mechanism that makes a speaker trustworthy in the way that we can open the back of a clock, to insist that testimonial trustworthiness must be historically grounded is like insisting that a clock could not count as repaired until it had run properly for a while. Just as an unreliable clock can be made reliable more or less at once by undergoing a repair of its

with Faulkner that one way is by manifesting trust in her. But what is the relation between S's reason to tell you the truth and the reason that you have, on your side, to believe what S tells you?

Say the intervention succeeds. Imagine that your trust succeeds in giving S a reason to tell you the truth and that S acts on that reason in telling you that p. Imagine further that S's testimony gives you an epistemic reason to believe that p. I agree with Faulkner that your trust can give S a reason to tell you the truth, but can it also inform the reason that you have to believe what S tells you? Here Faulkner and I disagree. Faulkner argues that because the way in which your trust gives S a reason to tell you the truth makes it more likely that S is telling you the truth, your trust can count as evidence that S is telling the truth. I think that this argument proves too much, since it would let you bootstrap your way into possessing reasons that you clearly do not possess. What Faulkner's account leaves out is what you hope your trust will cause: here-and-now reliability in the speaker. Of course, Faulkner does not altogether overlook how the speaker's perhaps new-found reliability figures in your reason to believe what she tells you, since he emphasizes that your trust can make the speaker more likely than she would otherwise have been to tell you the truth. But he views the core of your reason to believe that p as deriving from the *reason* that you give S to be reliable in telling you that p, rather than from S's actual reliability.

It may at first seem that my objection addresses only a narrow issue in the epistemology of testimony: whether a reason to believe testimony can derive from the addressee's trust itself or only from reliability in the speaker that the trust perhaps causes. But beyond my narrow disagreement with Faulkner lie two broader issues. In section II, I'll argue that Faulkner misappropriates Bernard Williams's genealogy of testimony when he makes use of Williams's genealogical argument in his own preferred assurance view of testimony. Though Williams doesn't clearly articulate it, there is a deep reason why Williams's genealogy cannot underwrite an argument for trust-based testimonial reasons.

---

time-telling mechanism, so an untrustworthy speaker can become trustworthy more or less at once by undergoing a reform of her truth-telling mechanism – that is, of her dispositions to be accurate and sincere. The latter process is usually not as straightforward as the former, and determining whether it has occurred is not as simple as inspecting some gears. But there is no reason to doubt that such a process can occur and, as Faulkner emphasizes, some positive reason to be confident that it can occur through the addressee's own intervention. (I'll press this analogy further in section I.)

This raises a second issue: can a genealogical argument underwrite any version of the assurance view of testimonial reasons – that is, any view that like Faulkner’s emphasizes the second-personal trust relation between speaker and addressee? In section III, I’ll argue that the answer is yes. I’ll sketch an assurance view of testimonial reasons that rejects Faulkner’s thesis that such reasons could be grounded in trust.

On my alternative assurance view, a testimonial trust relation derives from the addressee’s eliciting and from the speaker’s exercising a species of reliability that brings the two of them into a second-personal relation of mutual recognition. This is not the truth-conducive reliability on which reliabilist arguments in epistemology typically focus. It is instead a knowledge-conducive reliability that pertains to the addressee’s epistemic entitlement to treat the evidence at hand as sufficient, in his particular context, for him to close doxastic deliberation by forming a belief. I’ll thus call it ‘closure-conducive’ reliability. I’ll argue that when we see how warranted testimonial belief requires that the speaker be both truth-conducively and closure-conducively reliable – where the latter requires that she be responsive to the addressee’s context-sensitive epistemic needs, beyond his broad need to believe the truth – we’ll see how there is an ineliminable role for assurance in the epistemology of testimony. And we’ll see how filling that role requires rethinking what would count as a vindicating genealogy of testimony.

### **I. Is it trust that grounds a testimonial reason, or the reliability that trust causes?**

The core of Faulkner’s argument lies in Chapter 6, where he explains how the presumption of trustworthiness informing your affective trust in S provides what I’ll call a testimonial reason: an epistemic reason to believe what S tells you.<sup>2</sup> This argument is not quite complete in Chapter 6, since a key part of it rests on a claim that he defends in Chapter 7. In the genealogical treatment of testimony offered in Chapter 7, Faulkner grounds his account of testimonial reasons in an observation about our contingent social condition: that

---

<sup>2</sup> Faulkner aims to explain not testimonial reasons in general but only the testimonial reasons that derive from assertions considered as tellings – that is, as addressed to the one who gets the reason. He argues that there can be testimonial reasons that do not derive from assertions considered as tellings – for example, the reason an overhearer gets to believe what S tells someone else. This distinction won’t matter for my purposes in this paper.

we have left the State of Nature via the intrinsic value that we place on sincerity and accuracy, the twin virtues of truthfulness. This yields an account of testimonial reasons that is neither reductive nor straightforwardly non-reductive. On Faulkner's account, testimonial reasons cannot be reduced to reasons derived from our predictive faculties, because the affective trust at the core of testimonial relations is crucially different from predictive species of trust. But his account is not straightforwardly non-reductive either, since it denies the principle of credulity (on which we can get a testimonial reason by accepting what the speaker asserts entirely credulously).

Faulkner thus aims to sidestep the dialectic that informs much recent work in the epistemology of testimony. But how does Faulkner intend his account to work within the dialectical context that motivates it? That dialectical context emphasizes what Faulkner calls *the problem of trust*, a problem that I'll examine in detail as we proceed. On Faulkner's most general formulation, it is the problem that testimony is a product of the speaker's communicative intentions and as such should not be treated on the model of a regularity in nature. On the assumptions (a) that the audience is specifically ignorant of what moves the speaker to communicate in a given case and (b) that the speaker may be motivated by self-interest, there can be no *general* norm entitling audiences to believe speakers. Faulkner's explanandum is the entitlement to believe not merely what the speaker says but specifically to *believe the speaker* – that is, to believe what she says on her say-so. When, then, is a given audience entitled to believe the speaker? How, that is, could the audience acquire a reason to believe on the speaker's say-so? Only, Faulkner argues, in a context of affective trust.

Faulkner's argument turns on a distinction between two species of trust, predictive and affective, which he defines as follows<sup>3</sup>:

(PT) A trusts S to  $\phi$  (in the predictive sense) iff (1) A depends on S  $\phi$ -ing, and (2) A expects S to  $\phi$  (where A expects this in the sense that A predicts that S will  $\phi$ ). (145)

---

<sup>3</sup> I've added the labels – '(PT)' and '(AT)' – but the definitions quote Faulkner's text verbatim.

- (AT) A trusts S to  $\phi$  (in the affective sense) iff (1) A depends on S  $\phi$ -ing, and (2) A expects (1) to motivate S to  $\phi$  (where A expects this in the sense that A expects it *of* S that S be moved by the reason to  $\phi$  given by (1)). (146)

These definitions yield two key differences between predictive and affective trust. The first difference concerns the nature of the expectation in clause (2) of each definition: the expectation in (PT) is purely predictive, whereas the expectation in (AT) is also normative – a matter of what A presumes that S *ought* to do. The second difference concerns the content of these expectations: in (PT) A expects that S will  $\phi$ , whereas in (AT) A expects that S will  $\phi$  for the reason that A is depending on her to  $\phi$ .

One might try to argue that the second difference entails the first. When you expect that S will  $\phi$  specifically for the reason that you are depending on her to  $\phi$ , does not your expectation presume that your dependence gives you a claim right over S, a presumption that in turn makes your expectation that S will  $\phi$  normative rather than merely predictive? No, your expectation that S will  $\phi$  for the reason that you are depending on her to  $\phi$  does not necessarily include a presumption that your dependence gives you a claim right over S. Sometimes it does, and those are the cases on which Faulkner is focusing. When you depend on S for the truth, it is plausible that your dependence gives you a claim right to the truth from S. But you might depend on S to  $\phi$ , while acknowledging that your dependence gives S a reason to  $\phi$ , given background conditions, without presuming that you have a claim right over S to  $\phi$  – that is, that you could rightly criticize or resent S for failing to  $\phi$ . Perhaps S is your student or research assistant, and you've asked her to do you a small favor in the context of your research. You might expect that your dependence on S gives S a reason to do you the favor, partly because one thing that S is in the business of doing is proving her research potential and will in turn depend on you to attest to that potential in a letter of recommendation. So you do expect S to do you this favor because you expect that your dependence on her to do it gives her a reason to do it. But you don't for a moment presume that you have a claim right over S for this favor. In fact, it's quite the opposite:

you're hesitant to impose upon S precisely because you expect her to be especially responsive to the imposition, given the institutional context in which you make it. Faulkner does not explicitly link the two components of (AT) in the way that I'm criticizing, so the point isn't an objection to his account. But the possibility of such a link will matter to an issue that I'm going to raise in section IIIb below.

How does Faulkner's appeal to affective trust solve the problem of trust, given that it does not falsify the two assumptions that generate the problem? Here is his core move:

[I]n trusting S to  $\phi$ , the grounds of A's attitude of trust are the belief that S can recognize his, A's, depending on S  $\phi$ -ing, and the presumption that this will move S to  $\phi$ . Thus, A will perceive the situation defined by this act of trust as one wherein S has a reason to  $\phi$ . So other things being equal A will presume that S will  $\phi$ . If this turns out to be true and S acts as A expects, S will have proved trustworthy. So in affectively trusting S to  $\phi$ , A presumes that S will prove trustworthy just as in predictively trusting S to  $\phi$ , A would believe this. This is not to suggest that trust involves A reasoning to this conclusion but is rather to claim that in trusting S to  $\phi$ , A makes this presumption. However, the presumption that S will  $\phi$  rationalizes A's act of trust in the same way that the belief that S will  $\phi$  would do so. Consequently, the act of trust is rationally *self-supporting* in that it is based on an attitude of trust, which through implying the presumption that the trusted is trustworthy, gives a reason for trusting. (151)

And here is the core move applied to the case of testimonial trust:

[T]he attitude of affectively trusting a speaker for the truth provides an epistemic reason for believing the speaker's testimony. For suppose A trusts S for the truth as to whether p and S tells A that p. Then A's attitude of trust, I argued [above], involves A accepting various propositions about S and the trust situation, where the acceptance of these propositions defines what it is to see depending on S for information as to whether p in the positive light of trust. So in affectively trusting S for the truth, A accepts that S will see his, A's, depending on S for information as to whether p as a reason to tell A the truth on this matter. So trust involves A accepting that S has a reason to tell him the truth, and accepting that S will act on this reason, other things being equal.... In accepting these things about S and the trust situation, A thereby presumes that S is trustworthy, or that S will tell him the truth and will do so for the reason that he, A, depends on S for this.

This presumption need not amount to the belief that S is trustworthy since its ground is things which need be merely accepted in the trust situation. However, this presumption, like the belief with the same content, makes it probable for A that *p* is true given that this is what S tells him. So A's attitude of trust raises the probability of *p*, when this is what S tells him. So A's trusting S for the truth, in a situation where S tells A that *p*, provides A with an epistemic reason to believe that *p*. (154)

Let's consider the core move in light of two worries. First, why are we reasoning from A's trusting presumption that S will see A's dependence as a reason to tell him the truth? The problem case is the one where S tells A that *p* *without* being motivated by any recognition of A's dependence – that is, where the presumption is false. Second, why not suppose that what needs to be shown is not that A's attitude of trust has given S a reason to be trustworthy but that A's attitude of trust *has succeeded in making* S trustworthy in this specific instance? In the kind of case we're considering, we're entitled to assume that S knows that *p*, and (if this is not entailed by the first assumption) that it is indeed the case that *p*. But we're not entitled to assume that S is relevantly trustworthy – since the problem of trust is precisely that S may be untrustworthy in a given case.<sup>4</sup>

Faulkner replies to the second worry that if A's attitude of trust can explain why S's telling is likely to be true, then that attitude of trust can itself figure as evidence. But, elaborating the second worry, we may note that the trust itself doesn't *directly* explain why the telling is likely to be true. What directly explains why the telling is likely to be true is the speaker's truth-conducive reliability – that is, her disposition to assert the truth – in this interlocutory context. The trust may explain why the speaker is reliable. But it's the reliability that explains why the telling is likely to be true. Compare: you can cause your clock to be a reliable indicator of the time by repairing it, but the repair job – even if it is ongoing: say you have to hold the cord at a precise angle to retain the electrical connection

---

<sup>4</sup> The parallel formulations in an earlier paper – “What is Wrong with Lying?” (*Philosophy and Phenomenological Research* 75:3 (2007), 553) – made it seem that Faulkner holds that the epistemic reason is provided not by the audience's affective trust but by the *de facto* affective trust relation between speaker and audience. The reason is created “when things go right,” “when this presumption is fulfilled.” I take it that Faulkner would no longer put his view like that. If there no reason unless the speaker is (or shows herself to be) as the audience's presumption represents her as being, then it is not the *attitude* of affective trust that provides the reason but the *relation* of affective trust, which is something else entirely: the attitude plus appropriate responsiveness to the attitude on the speaker's side.

– isn't itself evidence of the time. It causes evidence of the current time to be produced, but it itself is not that evidence. The evidence is what the clock says, given that the clock is reliable.

Faulkner would reject this comparison because trusting a clock for the time manifests predictive trust, whereas trusting a speaker for the time typically manifests affective trust, where affective trust differs from predictive trust insofar as its expectation of performance is normative – not merely an expectation-*that*, but an expectation-*of*. But why should this difference matter to the point at issue? With a little science fiction, we could imagine a multi-step process of creature construction moving from merely predictive to fully normative reliance: (a) a regular clock that works only if you hold the cord just so; (b) a speech-interpreting clock that works only if you keep saying 'Hold that connection'; (c) a mind-interpreting clock that works only if it interprets you as wanting it to hold that connection; (d) a norm-sensitive clock that works only insofar as it is responsive to your stance of normatively expecting it to hold that connection.<sup>5</sup> Your intervention or influence clearly does not count as evidence of the time in (a) or (b). But if your intervention or influence counts as evidence in (d), as it seems Faulkner would have to say (consistently with his theory, assuming that relevant background conditions are met), then why not say the same of (c)? But (c) seems a mere extension of (a) and (b). Why should the introduction of the normative element in (d) make this difference?

Faulkner's full answer to this question will rest on the genealogical argument that he develops in Chapter 7 and that we'll consider in section II below. But we can anticipate one issue for that argument by considering norm-sensitivity that is purely instrumental. You can 'give S a reason' to tell you the truth by betting her or by threatening her. In such a case, it may be that the bet or threat 'explains why' S is likely to be asserting the truth – because without the bet or threat S would have lied or been less careful. While you might cite the bet or threat in explaining why you're entitled to believe what S asserted, it seems very odd to say that the bet or threat itself figures as evidence or as the basis of an epistemic reason – *unless*, of course, we regard the evidence or reason as lying in S's status

---

<sup>5</sup> For the idea of a 'creature construction' see Paul Grice, "Method in Philosophical Psychology," in his *The Conception of Value* (Oxford: Oxford University Press, 1991), especially section V.

as a reliable testifier and regard that status as simply including the bet or threat. So what would be reliable is: S-having-bet or S-under-threat. And we can now perhaps make a move in the direction of Faulkner's conclusion. We can say that there's no need for a hyphenated condition when the explanatory factor gives S an intrinsic reason: we don't need to say 'S-insofar-as-she-cares-about-the-norms-of-trust' because her care about the norms of trust somehow figures in her practical identity. But this move presupposes a reliabilist framework that Faulkner wants to reject.

What Faulkner wants to say is that the bet or threat is *the* explanation why S's assertion is likely to be true, just as trust is *the* explanation in the central trust cases. But without an appeal to S's reliability, an account of A's testimonial reason would admit the possibility that A has bootstrapped his way into possession of a reason through his mere affective trust in S. The problem for Faulkner is that we do not think such bootstrapping is actually possible.<sup>6</sup> Trusting testimony no more generates a reason to believe what the speaker asserts than trusting a promise on its own generates a reason to perform acts that depend on the assumption that the promisor will keep her promise. In each case, I would argue, the trust can at best cause the speaker to be relevantly reliable, where the reliability would in turn provide the reason.<sup>7</sup> We can cast the threat of illicit bootstrapping in terms that parallel Faulkner's own problem of trust. The problem is this: that even though the

---

<sup>6</sup> Faulkner raises and responds to a worry about bootstrapping in the following passage, but this is not the worry that I'm pressing:

It might seem odd that trust can bootstrap itself into reasonableness in this way. However, this oddness should be lessened once it is clear that trust is both an attitude and an action and that what is being offered is an account of the interaction between these two aspects of trust. The dynamic by means of which reasons for trusting are generated can then be clarified by separating out the temporal stages wherein an act of trust follows a decision to trust. (151)

Here and in his discussion through 153, Faulkner seems to assume that the only worry about bootstrapping to which his account might be susceptible is the worry that an attitude of trust commits the trusting to act in a way that is insensitive to evidence that the trusted is not worthy of the trust. That is not my worry. I take for granted Faulkner's point that "in deciding to trust S to  $\phi$ , A does not decide to trust come what may" (152). My worry applies most sharply when there is no evidence of S's untrustworthiness available to A, though S is nonetheless unreliable. Faulkner claims that in such a case A may have an epistemic reason to believe what S tells him. I'm arguing that that would involve illicit bootstrapping.

<sup>7</sup> I lack space to defend the claim about promising here. For a full defense, see my "'You May Rest Assured': A Theory of Normative Powers," in preparation. Of course, promissory reliability is not the same as testimonial reliability.

presumption at the core of A's affective trust *can* make S trustworthy, *for all A knows* S is not in fact trustworthy in this instance – indeed, for all A knows, S is merely *exploiting* the presumption at the core of the attitude that she invites A to take toward her when she tells him that p. If we say that A's affective trust gives A a reason to believe what S tells him, we appear to be merely wishing this problem away. A's trust may include a *belief* that A has a reason to believe what S tells him, but that belief merely sets the stage for the problem of trust. If your attitude of trust itself gave you a reason to believe what the trusted speaker tells you, there simply would be no such problem as the problem of trust that Faulkner describes. The problem of trust is that the attitude of trust cannot itself provide a reason to believe the trusted, independently of the speaker's status as reliable.

Why does Faulkner eschew such an appeal to reliability? He regards a reliabilist framework as failing to do justice to the interpersonal element in testimonial trust relations. But he thereby overlooks the possibility of arguing as I will, with a emphasis on closure-conducive reliability. In order to make available a reason to believe what she says, a speaker must be both truth-conducively and closure-conducively reliable. The dimension of closure-conducive reliability gets the second-personally normative element fully in play, I'll argue, by ensuring that S's status as fully reliable includes her responsiveness to A's doxastic predicament. Since that's what serves to distinguish the addressee's stance from an overhearer's stance – that the former but not the latter trusts in a way that presupposes that he is the recipient of such second-personal responsiveness – I'll conclude that testimonial reasons are irreducibly second-personal, providing the epistemic upshot that we would expect to derive from an assurance view of testimony, on which a testimonial telling is an invitation to trust.

## II. Can a genealogy vindicate trust-based reasons?

I haven't thus far considered what Faulkner regards as the key piece of his argument for trust-based testimonial reasons: the genealogical argument that he develops in Chapter 7. Building on Bernard Williams's genealogical argument in *Truth and Truthfulness*, as well as on his own argument in Chapter 6, Faulkner now argues that the problem of trust can be disarmed when certain social conditions are met. Again, the problem of trust is that a given

speaker could prove unmoved by the reason that your trust presumes she has to tell you the truth. Faulkner argues that this possibility does not constitute a deep threat to the possibility of trust-based testimonial reasons because even a speaker unmoved by it will – assuming the social conditions are met – nonetheless have this reason to be truthful.

In the present section I'll argue that there are two problems with Faulkner's genealogical argument. First, Faulkner's argument seems to presuppose without argument the denial of a plausible and widely accepted internalist thesis about reasons. Second, his account cannot distinguish between the reason-giving force of affective trust and the reason-giving force of what I'll call 'institutional' trust, despite the fact that institutional trust cannot serve as the ground of a testimonial reason.

### **IIa. The problem of internalism**

While a speaker could prove unmoved by your trust in a given case, Faulkner's genealogical argument aims to ensure that the speaker does nonetheless have a reason to be moved by it. Let's call that reason her *aretaic* reason, to contrast it with the *testimonial* reason that your trust presumes that you have to believe what she tells you: 'aretaic' because it is a reason for her to speak with sincerity and accuracy, the virtues of truthfulness. For Faulkner's account of testimonial reasons to work, something in the broader social practice must ensure that the speaker does actually have the aretaic reason. Faulkner's strategy is to turn that necessary condition on trust-based testimonial reasons into a sufficient condition, arguing that if we have escaped the state of nature to the extent that we are entitled to take for granted that our affective trust gives any speaker who addresses us an aretaic reason to tell us the truth, then we are equally entitled – epistemically entitled – to believe what the speaker asserts simply on her say-so. If the argument works then the problem of trust is not the problem that it appeared to be – at least, not for us. More exactly: the problem of trust becomes the problem that a speaker may prove unmoved by a reason to tell you the truth that she does nonetheless continue to have. As we'll see, this domesticates the problem of trust by delinking it from a natural application of an internalist thesis about reasons: that in order to have a reason to  $\phi$ , S must

have at least a sound deliberative route to a motive to  $\phi$ .<sup>8</sup> Faulkner argues that that element of risk is compatible with our nonetheless having a testimonial reason. After all, even accounts that put S's reliability at the core of A's reason to believe what S asserts have to live with the possibility that S's assertion will prove false though S herself is reliable. Faulkner aims to relocate the risk from A's reliance on a belief-forming mechanism presumed reliable to A's trust in a intentional agent presumed to be moved by that trust. Doesn't the latter formulation better capture the riskiness of taking someone at her word? Indeed it does. I'll endorse an alternative formulation of that riskiness when I sketch my alternative account of testimonial reasons in section III.

We're now ready to see the first problem with Faulkner's genealogical argument. Why should we believe that the aretaic reason at issue – that is, the speaker's reason to be truthful, grounded in your trusting dependence on her to be truthful – is a reason that just any speaker in our practice will have? Perhaps if the speaker is 'one of us' in relevant respects she must value sincerity – construed as involving the dependence-responsiveness posited by affective trust – *in general*. But it would be an absurdly strong claim to say that a given speaker must, to count as a participant in our practice, prove responsive to *your* dependence on her. The protagonist in the problem case is someone who may value sincerity in general but who in a given case *makes an exception of herself*. She wouldn't lie to her friends but she will to you, her mere business partner. Or she wouldn't lie to people whom she believes likely to smoke her out, but she will to you, whom she believes naive. Perhaps she wouldn't lie to anyone in her quotidian sunny mood, but that changes with the descent of crepuscular anger. And so on. Can genealogical reflections show that affective trust gives rise on its own to any reason for an audience to believe what a speaker tells them – even in social contexts where affective trust is generally valued? The problem of trust seems undomesticated.

Faulkner claims that the normative expectation at the core of affective trust is reason-giving at least in communities bound together by norms of trust. In such a

---

<sup>8</sup> For this use of 'sound deliberative route' see Williams's "Internal Reasons and the Obscurity of Blame," in *Making Sense of Humanity* (Cambridge: Cambridge University Press, 1995), 35; which extends the argument in "Internal and External Reasons," in *Moral Luck* (Cambridge: Cambridge University Press, 1981).

community, failure to live up to this expectation brings a sanction administered through reactive attitudes. This gives anyone who is committed to the norms governing community interactions – anyone who thereby gives standing to the reactive attitudes – a reason to live up to the expectations. And perhaps (though Faulkner doesn't argue this way) only an insider, someone who shares the identity, can understand how the reactive attitudes are to be administered and thus what motives they inculcate. So perhaps – if the identity is somehow constructed or sustained by that encounter – we can get round the worry that the reason-giving aspect of the trust is merely predictive. But now the question, the whole problem of trust, simply shifts ground: why think this speaker is committed to these norms *in application to the present case*? To say that she is committed to the norms in application to this case is to say she is relevantly trustworthy. So in wondering whether she is committed to the norms we're back wondering how to solve the problem of trust. The loop through local identities seems not to have helped.

The problem again appears to derive from an assumption that Faulkner does not make explicit: the falsity of a practical internalism about such aretaic reasons. On this internalism, there's a link between the speaker's possessing the reason that your presumption, when you trust, presents her as having and some fact about her motivations – let's call it, generically, the fact that she *could care* about not misleading you, whether for your own sake or in order to avoid others' disapprobation. The problem of trust is that *for all you know* the speaker not only does not care about not misleading you but could not be brought to care. I don't think we need to appeal to the concept of a psychopath to make this possibility clear. There may be people who simply dislike you or are angry with you to such a degree that they are incapable of caring not to mislead you, and there may be predicaments that make people desperate or despairing to such a degree that they are incapable of caring not to mislead any addressee, however they may feel about him. But it's easier to focus on psychopaths, and to note that a psychopath is defined as someone who could not care – someone who has a deficit in their capacity to care – about the normative pressures that Williams's and Faulkner's genealogical stories emphasize. If internalism is true, the problem of trust is that for all you know the speaker addressing you is a psychopath and as such does not have a trust-based reason not to mislead you.

Of course, Faulkner can avoid the problem by simply taking for granted that internalism is false. Indeed, that seems to be his strategy. But the denial of internalism – that is, externalism – is a highly controversial position about the nature of aretaic reasons, just as it would be about the nature of practical reasons in general.<sup>9</sup> And the problem for Faulkner is actually worse than that, since his problem of trust would appear to survive any such assumption. If we assume that even a psychopath has a reason not to mislead you, however incapable she is of being motivated to act on the reason, then the problem of trust becomes the problem that for all you know the speaker is incapable of being thus motivated. The fact that any speaker does, assuming externalism, have an aretaic reason to be truthful would not at all tend to show that you have a testimonial reason to believe what she says.

These reflections on internalism and externalism reveal an incoherence in the dialectic framing Faulkner's argument for trust-based reasons. His solution to the problem of trust – the solution that works by positing a link between the speaker's aretaic reason and the addressee's testimonial reason – actually depends on rejecting an externalist view of aretaic reasons. If we adopt an externalist view of aretaic reasons, then showing that the speaker possesses a reason to be truthful does not solve the problem of trust, since on an externalist view of reasons a speaker can have a reason to  $\phi$  while lacking a sound deliberative route to a motive to  $\phi$ . On an externalist view of reasons, a speaker can thus possess a reason to be truthful while being, as we might now put it, *deeply unmotivated* to tell her interlocutor the truth. This possibility shows that if he adopts or assumes an externalist view of aretaic reasons then Faulkner cannot solve his own problem of trust. If Faulkner instead embraces an internalist view of aretaic reasons, he thereby gets round this problem, since an internalist view ensures that a speaker who has a reason to tell her interlocutor the truth cannot be deeply unmotivated to tell him the truth. But now Faulkner confronts directly what I've been calling the problem of internalism: an internalist link between having a reason to be truthful and being motivated (or having a sound deliberative route to a motive) to be truthful ensures that speakers with psychopathic tendencies do not

---

<sup>9</sup> For a treatment of the debate over internal and external reasons that emphasizes interpersonal trust in practical reason-giving, see my "Trust and Reasons," in preparation.

have a reason to be truthful. Whichever view of aretaic reasons Faulkner adopts, his argument confronts a serious problem. The dialectic within which he is pursuing his argument for trust-based testimonial reasons appears to have painted him into a corner.

### **IIIb. The problem of institutional trust**

Let's now put that first problem aside. Even if we imagine that Faulkner's genealogical argument could somehow ensure that you'd never receive testimony from a psychopathic or otherwise similarly incapacitated speaker, it nonetheless confronts a second problem. Recall that by Faulkner's definitions (PT) and (AT) – given in section I above – affective trust differs from predictive trust in two respects: in its nature, affective trust involves an *expectation of* the trusted, not merely an *expectation that* the trusted will act as she is trusted to act; and in its content, affective trust specifically involves an expectation of the trusted to act on a motive generated by recognition of the trusting's dependence on her. The problem is that there are cases that satisfy Faulkner's definition of affective trust, (AT), without thereby amounting to cases that are *second-personal* in the way that Faulkner assumes any case of affective trust must be. The problem cases are cases of institutional trust, wherein being trustworthy in relevant respects is part of the trusted's role or job within the institution. The institutional nature of the trust ensures that it differs from predictive trust in both of the respects described by (AT). But the institutional nature of the trust equally ensures that it is not second-personal and therefore that it does not involve trust-based reasons.

Cases of institutional trust are common. I trust the bank teller to give me prompt and competent service concerning my banking needs, and in doing so I expect her to be moved by my dependence on her but not solely or even primarily by my *particular* dependence on her. I merely trust her to do her job. Part of doing her job involves knowing when and how customers are depending on her in ways to which her job requires her to be responsive. Say I know that this teller hates me and would probably mess with me in another context. Say I also know that, like most people, she takes her job seriously and acts with the integrity that is the natural expression of that attitude. When I trust her to transfer or deposit funds, my trust manifests a normative presumption not about her relation

to me but about her relation to her job and more broadly to the institutional context in which she performs it. That presumption is indeed fully normative: I expect no less of her. And it does involve the thought that she is motivated by my dependence on her. If she failed to be responsive to a customer's banking needs, I would judge her harshly, and expect others to judge her harshly, with the full reactive-attitudinal wallop characteristic of such judgments. (I expect no less of them, and I expect no less of myself than to accord others the respect that such normative expectations express.) I am not merely predicting that she's reliable. I may have no basis for such a prediction. If the little that I know about her suggests that I must have some predictive expectations, switch to a case in which I trust a store clerk about whom I know nothing. Barring a reason for mistrust, I'll trust the clerk to meet my pertinent shopping needs, on no other basis than that it is part of his job. Again, it isn't about his attitude specifically toward *me*. He may barely have seen me or may address me as part of an amorphous crowd of shoppers. Though these details do not entail that testimonial reliance on the teller or clerk could not be predictive – with added details, it appropriately could be predictive – these do not sound like cases in which trust in the speaker is affective in Faulkner's sense, with emphasis falling on a presumption about the speaker's attitude toward this specific instance of trusting dependence. In each case the trust seems fundamentally institutional, with emphasis falling on the relation not between the trusted and the trusting but between the trusted and her job or other institutional affiliation.

I agree with Faulkner that an assurance view of testimony ought to emphasize the second-personal trust relation specifically between the speaker and the addressee, not a species of trustworthiness grounded in the speaker's relation to her job or other institutional affiliation. But how in general should we think of the difference between institutional trust and a genuinely second-personal species of trust? Here again I think it helps to use the fiction of a creature construction; we might thus imagine a transition from a merely predictive reliance on a bridge through institutional trust in someone doing his best to replicate a bridge to, finally, genuinely second-personal trust in such a bridge replicator. Consider four cases (and forgive the cartoonish nature of the final three; I'm aiming at simplicity): (a) you rely on a short but rickety bridge to carry you across a dangerous chasm (just a bit too far to jump), (b) you rely in the same way on a rigid-looking man whom you

encounter bizarrely but apparently securely spanning the chasm (his eyes staring vacantly, his toes and fingers gripping gnarly roots), (c) you rely on a man in that pose whom you know to have been hired by the forest service to perform this important job but who is otherwise inattentive (reading a magazine, humming to himself, etc.), (d) you rely on a man who may or may not be making any money in the chasm-spanning racket but who looks you in the eye and asks you to trust him.

Since there are no expectations-of in play in (b), I expect Faulkner to share my sense that that case is no more an instance of second-personal trust than (a). And I expect Faulkner to share my sense that (d) clearly is an instance of second-personal trust. But what of (c)? It seems clear to me that if (b) isn't second-personal trust – that is, the kind that figures in the species of trust relation that Faulkner is trying to theorize – then neither is (c), even though (c) generates a expectation of performance that seems to meet clause (2) of Faulkner's definition of affective trust, (AT). What generates the expectation in (c) are the considerations (i) that serving as a reliable bridge is this man's job (with all that that involves) and (ii) that this particular job has moral implications. (Note that I'm not asking whether the man has a moral obligation to serve as a bridge. I'm asking whether your reliance on him as a bridge has moral content. In case (b) it does not, but in case (c) it does: you're counting on him to take seriously his important job.) If the man in (c) fails to do his job, that may naturally trigger a reactive-attitudinal response.<sup>10</sup> And a crucial part of what it is for him to do his job is to be motivated appropriately by the recognition that a traveler is depending on him to avoid plunging to his or her death.

As far as we've described case (c), there is nothing to show that you are not just predictively trusting this man, albeit in a way that has moral, and *thereby* reactive-attitudinal, content. For one thing, your moral expectation has nothing specifically to do with his relation *to you* – toward whom, after all, he is being quite inattentive. (Imagine again a rude clerk: you suspect he would gladly steer you in specifically the wrong

---

<sup>10</sup> I'm not sure that *resentment* is the right term for your reactive attitude here. Even though Strawson used 'resentment' as the catch-all term in his famous paper, it seems likely that you'd not be so much resentful as *disappointed* in the man, a disappointment that has nothing specifically to do with his relation *to you* (apart, of course, from the fact that he has let you fall into the chasm! – but that would be true in cases (a) and (b) as well, should they go wrong).

direction if his job didn't demand otherwise. You have expectations of him, but this is not second-personal trust.) From your perspective, you assess his bridge-relevant reliability in generically the same way as you do in (b). The specific difference is that in (b) you're interested only in organismic qualities of the man (is he asleep? in a trance? stiff-enough?), whereas in (c) you're also interested in some social qualities (does he take his job seriously? is he well enough paid?). But in each case you're merely looking for evidence of an incipient unfortunate collapse. What makes (d) crucially different in this dimension, I would say, is that your reliance on him naturally rests on a presumption of trustworthiness that projects a relation of mutual recognition. But if that's right, then the key to distinguishing non-second-personal trust – whether predictive or institutional – from genuinely second-personal trust lies in grasping the force of a species of mutual recognition that can't be defined along the lines of Faulkner's (AT). We'll return to this idea in section III.

Let's see how the problem arises for Faulkner's (AT). In section I, I suggested a link between the two components of affective trust, as Faulkner defines it in (AT). I suggested that when you expect that a speaker will  $\phi$  specifically for the reason that you are depending on her to  $\phi$ , your expectation presumes that you have a claim right over her, which in turn makes your expectation normative rather than merely predictive. I went on to criticize the suggestion, arguing that what amount to cases of institutional trust – in my example, trusting a research assistant to help you with your research – may manifest an expectation that S will  $\phi$  for the reason that you're depending on her to  $\phi$  without thereby manifesting any presumption that you have a claim right over S to  $\phi$ . But suppose the link nonetheless holds. We can now see that even such a link would not distinguish institutional from properly second personal trust. Even if your expectation that S will have a reason presumes that you have a claim right over S, that claim right may be mediated by an institutional context in a way that deprives it of second-personal content of a sort that could be articulated in terms of a trust-based reason. Even if your dependence on the teller gives you a claim right to prompt service from her, or if your dependence on the clerk gives you a claim right to conscientious advice about some product that she is trying to sell you, these claim rights are not themselves grounded in your trust but in the institution structuring the

exchange. I don't deny that there is a genuine phenomenon toward which Faulkner is gesturing with his definition of affective trust, (AT). But (AT) has not captured that phenomenon. The phenomenon in question is a trust relation that is irreducibly second-personal, involving a claim right that pertains specifically to the parties to this particular trust relation. As we'll see in section III, what (AT) leaves out is the way in which second-personal trust is mediated by a relation of mutual recognition.

One might at this point wonder if there is such a thing as a genuinely second-personal trust relation. If institutional trust can fail to be second-personal even when the trusted is motivated by recognition of the trusting's dependence on her, why think testimonial trust is ever second-personal? Here the dialectic gets more complex than I can cope with in this paper. The dialectic that we're pursuing begins from a point of agreement with Faulkner's argument: that, even if the epistemology of testimony is not simply the epistemology of testimonial tellings,<sup>11</sup> the latter is (a) a crucial part of the former and (b) a matter of irreducibly second-personal trust relations. My argument in the present subsection is that Faulkner is not in position to offer a compelling explanation why this is so, for the simple reason that his attempt to define the distinctively second-personal element in testimonial trust falls short of its aim, since it is compatible with cases in which the trust is not second-personal. One way to put my point is to note that Williams's treatment of testimonial trust emphasizes what we might naturally understand as its institutional nature – not that Williams conceives everyone as having a 'job' such that every addressee is in some respect a 'customer' or 'client' but that Williams conceives speakers as having been inculcated into a practice of what he calls 'normal trust,' a practice that has a fundamentally institutional nature insofar as it is upheld by norms of shaming, shunning, and the like. In this respect, as Faulkner acknowledges, Williams does not conceive of testimonial trust as purely predictive.<sup>12</sup> I'll discuss Williams's approach more fully in section III. My present point is merely that I agree with Faulkner that Williams's approach leaves out what

---

<sup>11</sup> See again note 2.

<sup>12</sup> On 174, Faulkner notes that Williams's solution to the problem of cooperation is not a reductive solution. I think this effectively concedes that Williams's conception of trust is not merely predictive. Though his account of testimonial reasons is reductive (as Faulkner rightly notes), Williams's conception of the assessment that that hearers must make of the speaker is not purely historical or predictive but typically refers to what I'm here characterizing as institutional elements.

Faulkner wants to emphasize: the second-personal nature of trust. So there is a burden on Faulkner's shoulders to explain how his approach differs from Williams's approach. But it does not appear that he can do that, given that his attempt to characterize the second-personal element in testimonial trust, (AT), admits cases that are purely institutional and as such lack the second-personal element.

One way to put this is to say that these cases of *merely* institutional trust do not involve an assurance. Of course, many cases of institutional trust do involve an assurance. Tellers and clerks often address you in the genuinely second-personal manner characteristic of a testimonial assurance. But sometimes they do not. Sometimes a teller or clerk merely asserts that p within your earshot, acknowledging that you are as a customer or client depending on her for the information whether p but without representing herself as aiming to do justice to your epistemic needs. I'll say much more about the relevant concept of an epistemic need in sub-section IIIb. My present point does not, however, depend on those details. It is easy to imagine (or remember) how a teller or clerk may manifest an institutional trustworthiness when she asserts that p without inviting your trust by manifesting appropriate responsiveness to your epistemic needs as addressee. Perhaps she isn't looking at you. Perhaps she looks at you but with a bored expression that conveys no interest whatsoever in any doxastic-deliberative circumstance that might inform your need for the information whether p. Still, it's her job to tell you the truth whether p, and there's no reason to doubt that she is incompetent in her job or that she fails to take it seriously. Though I cannot offer a full defense of the claim here, I find it plausible that such a speaker is fundamentally violating a constitutive illocutionary norm on testimonial tellings.<sup>13</sup> By a 'fundamental' violation I mean what J. L. Austin called a 'misfire,' not merely an 'abuse.'<sup>14</sup> It is an illocutionary abuse when S fails to do justice to A's epistemic needs in ways relevant to her testimonial telling. But there are cases in which S does not present herself as even attending to, or as trying to do justice to, A's epistemic needs. These are cases in which S's assertion manifests an aim of ignoring or dismissing A: she asserts that p within A's earshot but without addressing A in the way that Faulkner and I would agree is required

<sup>13</sup> For a full defense of this claim, see my "Assurance and Warrant" (*Philosophers' Imprint*, forthcoming).

<sup>14</sup> For the distinction between these two forms of illocutionary 'unhappiness,' see J. L. Austin, *How To Do Things With Words*, second edition (Cambridge: Harvard University Press, 1975), 13-17.

for her speech act to count as telling him that *p*. My argument in this sub-section is that the institutional context may ensure that the terms of (AT) are satisfied, simply because *S*'s failure to address *A* in a genuinely second-personal way is compatible with *S*'s being aware that *A* is depending on her for the truth, where providing such an auditor with the truth is required by *S*'s job, which she takes seriously. By (AT), such a case should – implausibly – count as a testimonial telling. And by Faulkner's broader argument, *A* may – absurdly – count as thereby acquiring a trust-based reason. These results are unacceptable by the lights of what I agree with Faulkner is the core insight of the assurance-theoretic approach to testimony.

Note well that neither the problem of internalism nor the problem of institutional trust is a problem for Williams's genealogical argument in the context of his reductionist approach to testimony. Williams developed and defended an influential version of internalism about practical reasons, and the species of trust that his genealogy vindicates as generally reasonable is primarily institutional, emphasizing not one-to-one trust relations but more general practices of inculcating and valuing the virtues of truthfulness. Each problem arises for Faulkner because of a key respect in which his position or argument differs from Williams's. Faulkner rejects Williams's reductionism in favor of a version of the assurance view on which affective trust can give reasons that are not grounded in the speaker's status as relevantly reliable. But (a) he assumes without defense an externalist hypothesis about aretaic reasons required to avoid the problem presented by psychopathic (or otherwise deeply uncaring) speakers, and (b) he does not seem to appreciate the possibility of a normatively robust institutional species of trust that is not yet affective, in his sense. These differences indicate key points of contrast between Williams's genealogical argument and Faulkner's redeployment of it. The argument of Chapter 6 for trust-based reasons has yet to receive its genealogical vindication. Faulkner's defense of an assurance view of testimonial reasons remains incomplete.

### **III. Can a genealogy vindicate second-personal reasons?**

Does an assurance view of testimonial reasons need genealogical vindication? I'll now argue that an assurance view needs only as much genealogical vindication as Williams

gives his reductionist view of testimonial reasons. Faulkner argues that Williams's genealogical argument is in a key respect incompatible with Williams's own emphasis on truthfulness as a species of trustworthiness. I think that misses the point of a genealogical vindication of truthfulness as a species of trustworthiness. Faulkner is wrong to claim that the point of a genealogy is to vindicate the possibility of trust-based reasons. The point of a genealogy is to vindicate, if we can, the intelligibility and real-world applicability of the concepts of reliability and trustworthiness that we use to explain the possibility of reasons that employ those concepts. The key to defending an assurance view of testimonial reasons therefore lies in showing how the concept of a testimonial assurance is informed by concepts of reliability and trustworthiness that can receive such a vindication. We can defend an assurance view without positing trust-based reasons but not without a vindication of relevant concepts of reliability and trustworthiness.

### **IIIa. How an assurance view needs genealogical vindication**

To see this, let's step back and ask what a genealogical argument might do for a theory of testimony. A genealogical argument addresses a question about obligation, in a way designed to engage skepticism about obligation. Does one really have an obligation to keep one's promises? Does one really have an obligation to tell the truth? Here are three answers: (a) yes, in all cases; (b) no, in no cases; (c) yes, in most (but perhaps not all) cases. A debunking genealogical argument explains why the fantasy of (a) obscures the truth of (b). A vindicating genealogical argument explains why the fantasy that the answer must be all-or-nothing – that is, (a) or (b) – obscures the truth of (c). The proponent of a vindicating genealogy, such as Williams, rejects (a) but aims to avoid (b) by vindicating (c).<sup>15</sup>

Why despair of (a)? Here again lies the problem of internalism. It's fine to say that one always has an obligation to keep a promise or to tell the truth, if what we mean is that not doing so amounts to a moral or illocutionary misstep. The question is whether one always has a reason to avoid such missteps. What if someone really did not care to step in line with the rest of us, regarding moral or illocutionary 'obligations' as fictions imposed

---

<sup>15</sup> The debunking versus vindicating contrast doesn't map neatly onto the contrast between Nietzsche and Hume. As Williams argues (*Truth and Truthfulness*, 12-19, 37-38), there are vindicating elements in Nietzsche – though he aims to debunk morality as *we* understand it.

by the powerful? What if someone were determined to avoid ‘buying into’ these fictions? Well if the person really lacks any sound deliberative route to a motive to uphold her moral or illocutionary obligations, then it is not absurd to wonder if she really has any reason to uphold them. If we conclude no, then she would have the obligations in one sense but not in another. Her conduct would be correctly criticized if she failed to keep a promise or to tell the truth, but not because we must view her as acting contrary to one of her reasons. A vindicating genealogy aims to show that we can be confident that such cases are rare because the normal conditions in which we make promises or give testimony are shaped by obligation-supporting institutions that we all – even the above-described renegades – have reasons to promote. We all – even the renegades – have instrumental reasons to promote these institutions, and the institutions in turn give anyone who is not such a renegade obligating reasons to keep a promise or to tell the truth.

Can Faulkner’s assurance view make use of such an argument? Faulkner plausibly criticizes Williams for leaving out the second-personal element in trust. But in endorsing Williams’s account of how we have escaped the state of nature in the respect relevant to testimonial reasons, Faulkner overlooks the deep connection between Williams’s vindicating genealogy and his implicit reductionism about testimonial reasons. As Faulkner observes, Williams treats testimonial warrant reductively, so Williams doesn’t need a genealogical vindication to do what Faulkner needs it to do. Williams needs to make sense of what it would be for a speaker to be generally reliable in testifying; appeal to the intrinsic values of sincerity and accuracy explains the nature of this reliability and thus what one would have to ascertain in assessing a speaker as reliable. Without the vindicating genealogy, Williams could not explain the terms in which we assess each other, and in which we expect to be assessed, when we give and receive testimony. But Faulkner’s argument imposes a heavier explanatory burden. Anyone who views testimonial reasons as grounded directly in trust must explain how it is that every speaker whom one might encounter, including the renegades, has a reason to tell one the truth.

As we saw in section II, Faulkner’s genealogical argument does not deliver this result. We can now explain that lapse by noting that that is not a result that a genealogical argument is designed to deliver. Faulkner appears to assume that the genealogical

vindication works for an entire practice – vindicating an obligation to keep one’s promise or to tell the truth for ‘our’ practice though perhaps not for other practices. But, as we’ve seen, a genealogy such as Williams’s does not aim to vindicate promissory or illocutionary obligations for the arbitrary speaker in a given practice. Williams’s genealogical argument explains why a typical speaker – not the same as an arbitrary speaker – has an obligation to be truthful. His genealogy does not show that every person has a reason to internalize the virtues, merely that there is a general reason to. Any given person can be an exception – gaining the cooperative goods without pulling her weight. Therein, of course, we confront the problem of trust. It is in recognition of this problem that Williams insists on his reductive approach to testimony. It is the same recognition that leads me to insist on grounding testimonial reasons in relevant reliability.

Williams does not explicitly appeal to an internalist thesis in the course of making his genealogical argument, so one might question my claim as an interpretation of Williams’s text. I do think that an internalist challenge is clearly enough figuring in the background of his argument.<sup>16</sup> But if one doubts that interpretive claim, there are explicit parts of Williams’s argument that entail the crucial difference from Faulkner that I’m emphasizing. Even if we set aside both of the problems discussed in section II – the problem of internalism, with its worry about psychopaths and the like, and the problem of institutional trust – we still get the problem that the addressee in a given case may not, as Williams puts it, *deserve* the truth. If you do not deserve the truth from a speaker, he argues, then that speaker has no obligation to tell you the truth. It is an important part of Williams’s genealogical argument that we not make a ‘fetish’ of assertion by regarding the illocutionary norm governing assertion as directly imposing an obligation of truthfulness – independently of whether the speaker’s addressee deserves the truth from her.<sup>17</sup> It directly follows that Williams is committed to denying that testimony *eo ipso* imposes an obligation

---

<sup>16</sup> Compare, for example, Williams’s endorsement of Nietzsche’s stance toward genealogy in *Truth and Truthfulness*, 12-19, 37-38, with his discussion of Nietzsche in “Nietzsche’s Minimal Moral Psychology,” (in *Making Sense of Humanity*, op. cit.); then compare the latter discussion with his treatment of internalism in other papers in that volume (cf. note 8 above).

<sup>17</sup> *Truth and Truthfulness*, 100-110. Faulkner endorses Williams’s underlying point (without the metaphor of a ‘fetish’) at *Knowledge on Trust*, 180-1. I discuss what Faulkner’s endorsement shows about how he conceives epistemic needs in section IIIb below.

of truthfulness. From the addressee's perspective, this is not merely an epistemic issue, whether you can know that the speaker addressing you is obligated to tell you the truth. The issue is whether you can have a trust-based reason to believe what she tells you – that is, a reason grounded in a reason that she might have to tell you the truth simply because your trust manifests dependence on her for the truth. You can have no such reason because she can have no such reason. If she does have a reason to tell you the truth, that's in part because you deserve the truth from her, and your status as deserving the truth from her is not determined by, and in fact typically has little to do with, your attitude of trusting her.

Williams does not aim, then, to show that every speaker has a reason to be truthful. This is why he appeals to the addressee's assessment to explain how the addressee gets a testimonial reason. But Faulkner is correct to note that such an appeal to assessment is incompatible with an assurance view, and it is plausible that Williams's emphasis on assessment is inconsistent, as Faulkner argues, with his conception of himself as in the business of vindicating norms of trust. Assessing the speaker for truthfulness is incompatible with simply trusting her,<sup>18</sup> and an emphasis on assessment erases the key distinction between believing a speaker and believing merely what she asserts. Since Faulkner wants to dispense with that need for assessment, he thinks he needs to show that every speaker has a reason to be truthful. And, as we've noted, it is hard to see how a genealogical argument could show that. This is the most fundamental problem confronting Faulkner's attempt to vindicate trust-based reasons. Is it fatal to any assurance-theoretic approach to testimonial reasons? How might an assurance view of testimonial reasons make better use of a genealogical argument?

### **IIIb. How an assurance view can ground reasons in reliability**

A genealogical argument for an assurance view need not show that every speaker has a reason to be truthful, because an assurance view need not embrace the idea that a

---

<sup>18</sup> In "Telling as Inviting to Trust" (*Philosophy and Phenomenological Research*, 70:3 (2005)) and elsewhere, I have argued that trust can be rational because it crucially rests on a counterfactual sensitivity to evidence of untrustworthiness: if there had been available evidence that the trusted is not worthy of your trust, you would not have trusted. Trust can thus be rational without the assessment for positive trustworthiness that Faulkner rightly regards as incompatible with trust.

testimonial reason can be grounded in trust. A testimonial reason cannot be grounded in trust, I've argued, because that would amount to illicit bootstrapping. The only way an assurance view can avoid illicit bootstrapping is to appeal to the speaker's status as relevantly reliable. But an emphasis on reliability need not be incompatible with an assurance view. Of course, if the testimonial reason rested entirely on the speaker's status as truth-conducively reliable, that would bypass the second-personal relation between speaker and addressee on which an assurance view focuses. But there is another dimension of reliability in play. To give a testimonial reason, a speaker must indeed be truth-conducively reliable. But she must also be appropriately responsive to how the quantity and quality of her evidence bears on her addressee's specific doxastic circumstances – that is, the context in which he would, if he trusted her, come to believe what she tells him. An emphasis on reliability can serve an assurance view if it targets the speaker's relation not only to the truth but also to her addressee's epistemic needs.

Like other proponents of the assurance view, Faulkner writes as if an addressee's epistemic needs are exhausted by his need to believe the truth. But that's too simple a conception of epistemic needs in general. When you wonder whether *p*, you're wondering not only whether it is true that *p* but whether you have sufficient epistemic warrant, in your actual circumstances, to believe that *p*. When you wonder whether to trust *S*'s testimony that *p*, you're wondering not only whether her assertion is true but whether it gives you, in your actual circumstances, sufficient warrant to believe that *p*. There are thus two burdens on the shoulders of any speaker who tells you that *p*: to tell you that *p* only when it is true that *p*, and to give you what would count, in your actual doxastic circumstances, as sufficient warrant for believing that *p*. As we'll see, we can easily imagine cases in which a speaker falls short of being a reliable testifier by doing the former but not the latter. That is, we can easily imagine how a speaker might tell you the truth without being appropriately responsive to your epistemic needs.

One key difference between my approach and Faulkner's thus derives from a difference in how we conceive the addressee's epistemic needs. Faulkner endorses the point that Williams emphasizes: that beyond telling what you believe to be the truth, sincerity requires doing so in a way that does justice to how much of the truth your

addressee needs and deserves from you.<sup>19</sup> We thus get Williams's case in which S is opening A's mail but when confronted tells A only that *someone* is opening his mail, not that she herself is. As Faulkner observes, S is here violating a crucial norm of trust – in the aretaic sense, S is being untruthful – despite telling A nothing but the truth.<sup>20</sup> Testimonial trustworthiness and the norm of sincerity require being appropriately informative in this respect: you must tell your addressee as much of the truth as he needs and deserves from you. That 'needs and deserves' marks a complexity that Williams discusses at length: A may make an unjust request for information that he does not deserve.<sup>21</sup> Whatever we say about that issue – the complexity that the norm of sincerity seems to rest on a norm of justice – it is undeniable that testimonial trustworthiness requires not merely that you tell your addressee the truth but that you more broadly do justice to his epistemic needs, giving him as much information as he deserves from you.

I accept that point, but my emphasis on the addressee's epistemic needs is different. When I note that a speaker might fail to be appropriately responsive to her addressee's epistemic needs, I mean that she might either tell him something that he doesn't, in his context, have sufficient epistemic reason to believe or fail to tell him something that he does in context have sufficient reason to believe (and that is supported by evidence available to S). It isn't at all difficult to come up with cases in which S has sufficient reason to believe that p in her context and does on that basis believe that p, but without being entitled to tell A that p – simply because it takes more evidence to count as sufficiently warranted to believe that p in his doxastic context than it does in hers. Allergy cases show this vividly: my pretty good evidence may suffice for me to believe that this bowl of snacks is nut-free but not for you to believe it, given your nut-allergy. I looked to see if the snacks contained nuts before I began to eat because I dislike the taste of nuts. "No nuts," I concluded, so I scooped up a handful. Now you arrive and ask me, "Does the bowl contain nuts?" I'm about to tell you that it does not contain nuts, since that's what I believe, but then I remember your allergy. "I can't say," I reply. Of course I *could* say, and with no impropriety – if I thought you merely shared my distaste for nuts. But your allergy

<sup>19</sup> Cf. note 17 above.

<sup>20</sup> Williams's mail-opening case is Faulkner's case 32.

<sup>21</sup> *Truth and Truthfulness*, 110-122.

imposes a higher standard on my telling. It would be a violation of illocutionary norms – of what Faulkner calls ‘norms of trust’; Austin called it an illocutionary ‘abuse’<sup>22</sup> – to treat you as entitled to believe by the evidence that suffices to entitle me to believe.

Though one may balk at calling a violation of illocutionary norms an ‘insincerity’ when the speaker does believe what she asserts, we may naturally call it a failure to be properly informative.<sup>23</sup> To get a case of this sort, let’s flip the previous case around and imagine that you’re the speaker, with a severe nut allergy, and I’m your addressee, known merely to dislike nuts. You’ve been checking out the snack bowl and are confident enough for my needs but not for your own that the bowl does not contain any nuts. I ask you if the bowl contains nuts, and you tell me that it does not – despite not yourself believing what you assert. Are you insincere? Are you in any respect attempting to deceive me? Well, we may imagine that your refusal to explain why you are not yourself eating from the bowl manifests an attempt to deceive me about your allergy. But that’s a different matter, and its relevance to the present issue is merely that it helps distinguish your illocutionary obligations from other aspects of your relationship with a given interlocutor. (Why, anyway, should I have a right to know about your medical status?) The topic of our actual conversation is this bowl of snack food, and you aren’t attempting to deceive me about *that*. If without explanation of your allergy (and, again, you may well not owe me any explanation) you refrained from telling me what you believe me entitled to believe in answer to a question I’ve just asked you, that *would* amount to an illocutionary violation – to an Austinian abuse – since it would count as withholding information that I manifestly need and deserve and that I am now explicitly requesting from you. But that is not the case we’re imagining.

Illocutionary norms require telling your addressee what he needs and deserves to know given *his* epistemic standard, the standard against which he would form a belief. (Of course, you use *your evidence*, not his. Obviously, the idea is not that you should tell him only what he’s in position to figure out on his own!) In cases where your doxastic circumstances differ, the epistemic standard that governs your testimony may well differ

---

<sup>22</sup> See note 14 above.

<sup>23</sup> In “Assertion, Sincerity, and Knowledge” (*Noûs*, forthcoming), I argue that *either* this is insincerity *or* sincerity does not mark an illocutionary norm.

from the standard against which you would form and retain a belief of your own. The norm governing testimony and the norm governing the speaker's belief are different norms that may give the speaker different directives: either 'tell that p but don't believe that p,' or 'believe that p but don't tell that p.' The norms can come apart in these ways because tellings are not announcements of what you believe but acts of assurance, wherein you put yourself under a norm defined from your addressee's perspective.

You must meet this norm – as we might put it, you must be *properly informative* – by being such that your addressee can rely on you as an informant. Flipping the testimonial relation around, you can get a reason to believe a speaker only if the speaker is a reliable informant – that is, reliable not merely as a speaker of truth but as a giver of such a context-sensitive assurance. The core of an assurance view of testimonial reasons lies in how testimonial trust – the species of trust that you manifest when you believe the speaker, not merely what she asserts – presumes that the speaker is reliable in both of these dimensions. Testimonial trust presumes that the speaker is not only truth-conducively but also, as we might put it, closure-conducively reliable – 'closure-conducively' because a speaker acts as a genuine informant only when her assertion that p permits her addressee to close a deliberation whether p, or to treat this matter of possible deliberation as closed, simply by believing her – that is, by accepting her assertion on trust. To accept an assertion on trust, you cannot be in the business of assessing the speaker for reliability. As Faulkner rightly emphasizes, when you assess for reliability you violate the terms of the trust relation, effectively stepping outside any relation of trust and believing (or not) on the basis of your independent assessment of the evidence. When you assess for reliability, you may wind up believing what the speaker asserts, but you don't believe the speaker: you don't accept what she says on her say-so.<sup>24</sup> So an assurance view needs to be anti-reductionist and emphasize that the speaker's reliability figures not in the addressee's assessment but as a defeating condition on whether an addressee who trusts without thus assessing counts as acquiring a testimonial reason. Anti-reductionists who are not assurance-theorists emphasize such a defeating condition on truth-conducive reliability, without realizing that there is a second

---

<sup>24</sup> I emphasized this point at length in "Telling as Inviting to Trust," where I also emphasized the need for reliability.

dimension of reliability in play. When we place equal emphasis on closure-conducive reliability, we open the door to an assurance view of testimonial reasons that avoids the problem of illicit bootstrapping.

Is closure-conducive reliability second-personal? From the perspective of an assurance view, the problem with appeals to truth-conducive reliability is that a speaker's reliability as an asserter of truth is not second personal: when she tells A that p, S's reliability in asserting the truth on the question whether p has nothing directly to do with her relation to A. But now what of S's reliability in giving A a reason that could serve to close A's deliberation whether p? Perhaps there's a guru with a website who can determine what would be a deliberatively sufficient reason for this or that abstractly characterized doxastic context, including a context like A's. Say S consults this website before telling A that p, and that the consultation is what makes S closure-conducively reliable in telling A that p. Does that show that closure-conducive reliability is not second-personal? It does not. It may be that the guru's website explains how S *became* closure-conducively reliable, but what her closure-conducive reliability *is* in this case directly involves her relation to A. The observation applies a point that I emphasized in a different application in section I: the nature of a reason is one thing, how one became able to give that reason another. In section I, I argued that a second-personal relation, mediated by affective trust, might explain how S comes to be truth-conducively reliable in addressing A, where truth-conducive reliability is not itself second-personal. Here I'm arguing that something that is not second-personal, consulting a third-party's website, might explain how S comes to be reliable in a way that is nonetheless second-personal. However it is produced, closure-conducive reliability involves a second-personal responsiveness, not mere possession of information or just any ability to 'get it right.' The guru's website may give S information that in turn enables A to count as closure-conducively reliable in addressing A. But what it is for S to be thus reliable makes essential reference to S's relation to A. One might object that what matters is S's relation to the doxastic context that A happens to be in, a relation that doesn't itself essentially involve A. But that context is defined by A's actual epistemic needs. Those needs can be characterized abstractly, but S's reliability consists in her responsiveness to those needs conceived *as A's* – that is, as the needs of this particular person to whom she

addresses her speech act.

One might still object that even if S must be responsive to A's epistemic needs, the species of reliability in question is reliability in addressing the needs of anyone in a doxastic context relevantly similar to A's. One might think that closure-conducive reliability must, like truth-conducive reliability, be implicitly general – a matter of getting a kind of thing right. A truth-conducively reliable speaker is reliable in getting the proposition right – that is, in asserting that  $p$  only when  $p$  really is true. And a closure-conducively reliable speaker is reliable in getting her interlocutor's doxastic context right – that is, in assuring A that  $p$  only when A really is entitled to close doxastic deliberation with the belief that  $p$  (or to treat that deliberative matter as closed). Viewed from this angle, S's reliability in getting her interlocutor's context right is no more second-personal than her reliability in asserting the truth.

A full reply to this objection would take us quickly into deep issues in epistemology and the philosophy of mind, since I would argue by drawing a comparison between testimonial reliance and a single subject's self-reliance when the subject forms a judgment that  $p$ . In a fuller treatment, I would argue that your self-relations when you form a judgment crucially include a relation isomorphic to the relation in which you stand to a speaker when you depend on her status not merely as truth-conducively but also as closure-conducively reliable.<sup>25</sup> When that relation is realized intrapersonally it becomes clear that it is not a relation between the subject and a source of information, conceived as a guide to 'getting it right.' Of course, you do rely on yourself to 'get it right' when you form a judgment, but that is a question of truth-conducive reliability: you rely on your epistemic faculties to give you the truth. The question of closure-conducive reliability here is question of self-concern: is your disposition to treat the doxastic question whether  $p$  as settled by your evidence a disposition that does justice to your epistemic needs? The question is not, of course, how you 'feel' about yourself but whether you are actually meeting those needs. You typically do trust yourself in this way, relying on your status as closure-conducively reliable every time you form a belief. Such self-trust is required even

---

<sup>25</sup> For part of this argument, see my "Reflection, Disagreement, and Context," *American Philosophical Quarterly* 49:2 (2012). For another part, see my "Judging as Inviting Self-Trust," in preparation.

in simple cases of perception. When you form a belief you treat the disposition that gives content to the self-trust – the disposition to treat your present evidence as sufficing to settle some matter that you might have deliberated, or deliberated further – as manifesting appropriate self-concern, by which I simply mean: as adequate to your actual epistemic needs. In any given case, you might not have done so; you might have treated the disposition as manifesting a mistake or confusion about your epistemic needs – or, in the most interesting cases, as manifesting a kind self-sabotage. This important dimension of your self-relations reveals a striking parallel between interpersonal and intrapersonal trust. In each dimension of trust, you rely on the trusted both truth-conducively and closure-conducively. The latter species of reliance is second-personal, or directed in the way of a second-personal relation, even in the intrapersonal dimension. Whenever you rely on a source or ‘mechanism’ of belief formation closure-conducively, you treat it as adequately concerned or ‘caring’ for – that is, as taking care of – your actual context-sensitive epistemic needs.<sup>26</sup>

These remarks raise large issues. But setting aside that larger inquiry, we can note that the presumption of closure-conducive reliability figures at the core of testimonial trust. Whatever we say about the deeper role of trust in judgment, your trust in a speaker who tells you that *p* rests most fundamentally on your dependence on her as closure-conducively reliable. We can see this clearly when we ask what drives the distinction that assurance theorists emphasize, between believing the speaker and believing merely what she asserts. In each case, you depend on the speaker as truth-conducively reliable. The distinction emerges insofar as in the former case, but not in the latter, you depend on the speaker as closure-conducively reliable. To depend on a speaker as closure-conducively reliable is to grant her executive authority over your beliefs – to treat her assertion not as mere input to your independent deliberation whether *p* but as on its own settling whether *p*. When you merely believe what she asserts, you do not treat her speech act as settling whether *p*; but when you believe the speaker, you do. To treat her speech act as settling whether *p* is to treat the speaker as not only truth-conducively reliable but also as closure-conducively

---

<sup>26</sup> One might try to distinguish trustworthiness in testimony from trustworthiness in judgment as merely an application of the distinction between the other-regarding and the self-regarding. In “Assertion, Sincerity, and Knowledge” I argue that in this application that distinction is not so simple.

reliable on the question whether *p*. Unlike truth-conducive reliability, closure-conducive reliability goes right to the core of the testimonial truth relation.

#### **IV. Comparing the two versions of the assurance view**

Though I reject Faulkner's aim of vindicating trust-based reasons, I agree with his observations about the curiously self-reflexive nature of trust. When you trust a speaker, you expect her to acknowledge your dependence on her. When you invite such trust, by telling someone that *p*, you expect that he will acknowledge that you aren't merely asserting that *p* within his earshot but assuring him, whereby you expect him not merely to listen to you but to trust you. As Faulkner emphasizes, these are normative expectations: not merely something you expect him to do, but something you expect *of* him. I would explain the key normative expectation simply as the expectation that the speaker is closure-conducively reliable. The self-reflexive nature of testimonial trust derives from the fact that when you trust a speaker you are relying on her not merely to speak the truth but to do justice to your epistemic needs – that is, to your need to be warranted in closing deliberation, or in treating it as closed, in your particular circumstances. Faulkner is right that testimonial trust is self-reflexive but wrong about how.

I agree with Faulkner that testimonial tellings are structured by norms of trust, but I regard those norms as themselves structured by a mutual recognition between speaker and addressee. What makes the structure of recognition normatively engaging is the acknowledgment that the perspective on the other side is a perspective on, among other things, *one's own perspective*. The speaker imputes to the addressee a perspective from which he looks back on her perspective and trusts her. The addressee imputes to the speaker a perspective from which she looks back and either provides or fails to provide an adequate basis for that trust – a basis that would lie, as we've seen, in both truth-conducive and closure-conducive reliability. If his sensitivity to evidence of untrustworthiness is triggered, thereby alerting him to the absence of that basis, his refusal to trust constitutes a claim of right: 'Hands off, *you* don't get to define *my* epistemic needs.' If he does trust, he accepts the provision of care. Of course, he could instead believe what the speaker tells him without trusting her, by refusing the illocution but accepting the evidence it gives him.

But that would also amount to a refusal of the care. Putting the addressee's recognition as a question of the speaker's care and the speaker's recognition as a responsiveness to the addressee's needs captures the second-personal dynamic at stake in the speaker's status as closure-conducively reliable.

Here, once again, is where Faulkner and I part ways. From my alternative assurance view, I'd explain the divergence as follows: Faulkner overlooks how closure-conducive reliability informs the recognitional structure that in turn informs testimonial trust. If you fail to see how reliability can be closure-conducive as well as truth-conducive, and how both forms of reliability inform warranted belief, you'll have a hard time resisting the idea that any role for trust in testimonial reasons must posit that trust as the ground of the reasons. I've argued that the idea is an illusion. There is a key role for trust in a broadly reliabilist account of testimonial reasons: trust necessarily structures the recognitions whereby the speaker brings her reliability to bear on her addressee's epistemic needs. In the natural and compelling metaphor, she invites his trust. She invites his trust by presenting herself as reliable in a way that would give him a reason to believe what she asserts. One puzzle that informs Faulkner's book is how presenting oneself as reliable could amount to inviting trust. My solution is that presenting oneself as closure-conducively reliable amounts to providing, or at least to undertaking an illocutionary commitment to provide, a distinctively epistemic species of care.

Again we might make the point vivid with a creature construction. Consider three cases: (a) you encounter a sign warning you of danger further down the path you're traveling; (b) you encounter a man making assertions, though none addressed to you, about the danger down the path; (c) the man turns to you, looks you in the eye, and tells you of the danger. On my view, the man in (c) is doing something of an epistemic nature that the man in (b) simply is not doing: presenting himself as closure-conducively reliable about your epistemic needs. On Faulkner's view, the man in (c) does nothing of an epistemic nature that the man in (b) is not doing. Indeed, the only thing of an epistemic nature that either man does is already done by the sign in (a): each presents itself as truth-conducively reliable. On Faulkner's view, the extra epistemic work in (c) is done by your presumption, as addressee, that your trusting dependence gives this man a reason to be truth-conducively

reliable that the man in (b) does not possess. My argument against that move has been two-fold: (i) even if your trusting dependence does give the speaker a reason – and Faulkner hasn't argued that it would give someone with psychopathic tendencies a reason – that would merely be an observation about how the speaker came to be reliable; (ii) Faulkner has not explained how this mechanism differs from a merely institutional mechanism whereby the man in (b) might likewise have a reason to be truthful, since he may be aware of your dependence on him despite failing to address you. But we can now see a broader rationale for suspicion of Faulkner's entire approach. An assurance view of testimonial reasons ought to give an epistemic basis for distinguishing the speech act in (c) from the speech act in (b), since the former contains an assurance that the latter lacks. An assurance-theoretic approach to the epistemology of testimony ought to say how it makes an epistemic difference that the man in (c) looks you in the eye and invites your trust. I think this difference lies in how his invitation to trust purports to manifest closure-conducively reliability, a status that is itself directly epistemic. An assurance view of testimony needn't eschew reliability because a species of reliability serves to distinguish testimonial assurances.

Can we in general count on speakers to manifest the second-personal concern at the heart of closure-conducive reliability? We thus ask whether the norm informing the practice of testimonial assurance can be rationally sustained. It is not obvious that it can be rationally sustained. Perhaps Williams is right, and all we can rationally sustain is a practice that institutionalizes a need to assess speakers for reliability. If he's right, then the assurance view of testimony is fundamentally wrong. It would take a genealogical argument, or something like it, to decide this matter. It is, I think, undeniable that we treat others as if they could give us reasons grounded not merely in truth-conducive but also in closure-conducive reliability. It is undeniable that we often, if not always, treat each others' assertions as inviting trust. Is this a rationally sustainable practice? I have suggested that our interpersonal presumptions mirror intrapersonal presumptions that lie at the heart of individual judgment. This amounts to the suggestion that a debunking genealogy would overturn more than merely our testimonial presumptions. If we aren't testimonially trustworthy in this dimension, it's hard to see how we could be worthy of our

own trust even as we form our solitary judgments.<sup>27</sup> Again, the question isn't whether trust is risky. If there weren't a risk that the trusted would prove unworthy of your trust, your reliance wouldn't count as trust. The question is whether the practice and presumption are rationally sustainable – whether it makes sense in general to let our social transaction, and perhaps our self-transactions, rest with a counterfactual sensitivity to evidence of untrustworthiness, rather than working from a positive assessment of trustworthiness. No genealogist of trust has attempted such a vindication, and I'm not sure what it would take to pull one off. We need a better understanding of the risks of relying on a closure-conducive provision of care, and of how we might manage if we had to do without it.

**Edward S. Hinchman**

*University of Wisconsin-Milwaukee*

hinchman@uwm.edu

## References

- Austin, J.L. (1975) *How To Do Things With Words*. Cambridge, MA: Harvard University Press.
- Faulkner, P. (2007) 'What is Wrong with Lying?' *Philosophy and Phenomenological Research* 75 (3): 535-557.
- Grice, P. (1991) 'Method in Philosophical Psychology', in *The Conception of Value*, Oxford: Oxford University Press.
- Hinchman, E.S. (2003) 'Trust and Diachronic Agency', *Noûs* 37 (1):25–51.
- \_\_\_\_\_. (2005) 'Telling as Inviting to Trust', *Philosophy and Phenomenological Research* 70 (3): 562–587.

<sup>27</sup> Compare this with my conclusion in "Trust and Diachronic Agency" (*Noûs* 37:1 (2003)): "If no one is trustworthy, no one has a reason to follow through either on his or her intentions or on others' advice. If no one is trustworthy, diachronic agency – the life of practical reason – is simply impossible" (45). We need a vindicating genealogy – or some kind of vindicating account – in the practical case as much as in the epistemic.

- Hinchman, E.S. (2012) 'Reflection, Disagreement, and Context', *American Philosophical Quarterly* 49 (2):95-111.
- \_\_\_\_\_. (forthcoming) 'Assurance and Warrant', *Philosophers' Imprint*.
- \_\_\_\_\_. (forthcoming) 'Assertion, Sincerity, and Knowledge', *Noûs*.
- \_\_\_\_\_. (in preparation) 'You May Rest Assured: A Theory of Normative Powers'.
- \_\_\_\_\_. (in preparation) 'Trust and Reasons'.
- \_\_\_\_\_. (in preparation) 'Judging as Inviting Self-Trust'.
- Williams, B. (1981). "Internal and External Reasons", in *Moral Luck*, Cambridge: Cambridge University Press, pp.101-113.
- Williams, B. (1995a) 'Internal Reasons and the Obscurity of Blame', in *Making Sense of Humanity and Other Philosophical Papers*. Cambridge: Cambridge University Press, pp. 35–45.
- Williams, B. (1995b) 'Nietzsche's Minimalist Moral Psychology', in *Making Sense of Humanity and Other Philosophical Papers*. Cambridge: Cambridge University Press, pp. 65–76.
- Williams, B. (2002) *Truth and Truthfulness: An Essay in Genealogy*. Princeton: Princeton University Press.

## THE TRUST GAME AND THE TESTIMONY GAME

**Katherine Hawley**

Paul Faulkner has given us a very rich and complex book, one which will repay careful study both now and in years to come, for anyone interested in trust, testimony or epistemology more widely. Rather than attempting to summarise his arguments, or to grapple with them as a whole, I will focus here on Faulkner's development of his keystone 'problem of cooperation', which he generates through discussion of the 'Testimony Game', and its parallels with the 'Trust Game'.

### **1. The Trust Game**

The Trust Game is an experimental set-up which is used by economists and social psychologists to investigate our behaviour and choices under certain constrained circumstances. In the standard Trust Game, a first player is given £10, and can choose either to keep all of this, or else to transfer some or all of the £10 to a second player. Whatever the first player decides to transfer is quadrupled by intervention from the experimenter. For example, if the first player decides to keep £6 the second player receives £16. The second player can then opt either to keep all she receives, or else to send some or all of it to the first player.

This is a peculiarly artificial situation, where the stakes are low (no-one risks their own personal funds), and players are somewhat distanced from ordinary social norms (it's just a game, after all). But social scientists like it because it allows them to put numbers on people's behaviour, by recording how much is transferred in each direction, and then to see how these numbers vary as the experimental set-up is varied. For example, the Trust Game can be played as a one-off, or repeatedly between the same two players. The game can be used with players of different ages, or different nationalities, or different genders, either in matching pairs or across these categories. Players may be allowed to interact face-to-face, to see photographs of one another, to speak by phone, or to have no direct contact

whatsoever. Players may be told they are members of the same fraternity or club. Players may be given a whiff of oxytocin, the hormone sometimes known as the ‘cuddle chemical’. Players’ behaviour in the game may be compared with their responses to poll questions about trustworthiness in society, and so on.

The first puzzle about the Trust Game concerns the behaviour of the second player. Even when the game is played as a one-off, a proportion of second players choose to transfer some money to the first player. And this behaviour is hard to understand in terms of self-interest. Suppose you have £16. You can either take it home, or else give some of it to a stranger with whom you will never interact again. Why do anything other than keep the whole £16? What is motivating those second players who decide to transfer some of their cash?

The second puzzle concerns the behaviour of the first player. Even when the game is played as one-off, many first players choose to transfer some money to the second player. This suggests that either the first players are predicting that the second players will not be motivated by pure self-interest, or else that the first players themselves are not motivated by pure self-interest. You have £10, which you can either keep, or else share with a stranger who will have no obvious motivation to send you any money in return. Why do anything other than keep the whole £10?

Collectively, these puzzles raise both a normative and a non-normative issue. The non-normative issue concerns the empirical mismatch between the assumption that people are motivated by narrow self-interest, and expect others to be likewise motivated, and the actual behaviour of people who play these games. Some players do not behave as these assumptions would predict: what, then is their additional or alternative motivation? (We might address this question by exploring other set-ups, such as the dictator game, in which the first player simply chooses whether or not to transfer some money to the second player. Either way, that’s the end of the game. A substantial minority of first players transfer money even in the dictator game.)

The normative issue arises from the fact that the first player ends up better off (takes home more than £10) if she takes a risk on the second player, and the risk pays off. And of course the second player benefits from any degree of transfer from the first player. If the

first player takes a risk, then the experimenter injects more cash into the game, which can potentially benefit both players. So it seems intuitively that there's good reason to take this initial risk, which conflicts with the idea that there is no good reason to do this. What is the rational choice for the first player?

The difference between the non-normative and the normative puzzle can be brought out by considering other experimental situations in which people commonly make mistakes in reasoning or judgement. (One example: when a character, Linda, is described as active in radical student politics, people often judge that in later life it's more likely that she becomes a feminist bank teller than that she becomes a bank teller. Another example: confronted with two-sided cards, people are bad at judging which cards they need to turn over in order to test the rule that if a card has a vowel on one side, it has an even number on the other.) In such situations, there is a non-normative puzzle: why do people commonly make this mistake? But there is no corresponding normative puzzle, no temptation to think that in some sense there is a genuinely good reason to make the judgement in question.

For Faulkner, the Trust Game illustrates the 'problem of cooperation'. He says "For the investor [i.e. the first player] to be acting reasonably in making a transfer, he needs to think, for whatever reason, that the trustee [i.e. the second player] will make a back-transfer. This game then illustrates how cooperation can be problematic because it is arguable that we often lack grounds for thinking this, but make transfers nevertheless. We trust and yet appear to be unreasonable in doing so." (p.4)

Is this the normative or the non-normative issue? I think it's a bit of both. There's the empirical fact that many first players make a transfer (i.e. 'trust' the second player), which they would not do if they were motivated purely by narrow self-interest and assumed that the second player was likewise motivated (and they were able to reason through the consequences of this assumption). So there's the non-normative question of what motivates such first players. But then there's also the normative question of whether what motivates such first players does after all make their choice a reasonable one. Faulkner suggests that we cannot see such choices as reasonable unless the first player is motivated by the belief that the second player will cooperate (or at least by a belief that this is fairly likely), and that the first player has grounds for this belief.

## 2. The Testimony Game

Faulkner then argues that the Testimony Game is importantly analogous to the Trust Game. Here is a description of the Testimony Game which attempts to make that analogy as closely as possible. The first player is the audience, and the second player is the speaker, who asserts that  $p$ , let's say. The first player begins, not with £10, but with no opinion as to whether  $p$ . This neutrality is worth something epistemically, in that it is better than a false belief as to whether  $p$ . The first player then has the option of keeping this neutrality (i.e. ignoring what the speaker says), and ending the game, or else giving some credence to what the speaker says. If he gives some credence to what the speaker says, the speaker (the second player) then has two options. The first is to take the credence and walk away—that is, to take the benefits of being believed, but not to return the favour by being trustworthy. The second option is to take the credence, and return something of benefit to the first player, the audience, by being trustworthy.

In the Trust Game, the experimenter injects extra cash iff the first player decides to make a transfer, and this extra cash can benefit both parties iff the second player also makes a transfer. In the Testimony Game, the analogous idea must be that the benefit to the second player (the speaker) of being given credence is greater than the potential loss carried by the first player (the audience) in moving from neutrality to giving credence; if the second player (the speaker) responds by being trustworthy, then both parties end up better off than they would have been had the audience ignored the speaker. Generalising, we are all better off if as a rule audiences give credence and speakers are trustworthy.

Recall that there are two initial puzzles about the Trust Game. Why does the second player transfer cash, as opposed to taking the whole lot home, as narrow self-interest seems to dictate? And why does the first player transfer cash, given the assumption that the first player is motivated by narrow self-interest, and assumes that the second player is likewise motivated? These then generate a non-normative question: why do people behave in this way? And they generate a normative question: narrow self-interest seems to make the first player's transfer unreasonable, but the overall benefits of cooperation seem to make the first player's transfer reasonable. So which is it?

In the Testimony Game we may likewise ask why the speaker (the second player) decides to be trustworthy, as opposed to saying whatever suits her immediate self-interest best. And we may ask why the audience (the first player) offers credence, on the assumption that she is motivated by narrow self-interest, and assumes the speaker is likewise motivated. Then there is the non-normative puzzle: what motivates audiences to offer credence, and speakers to decide to be trustworthy? And the normative puzzle: narrow self-interest seems to make the audience's decision unreasonable, but the overall benefits of cooperation seem to make it reasonable. So is it reasonable or unreasonable?

Faulkner articulates the problem of cooperation in the context of the Testimony game as follows: "the acceptance of testimony must be backed by reasons if it is to be reasonable. The problem of cooperation is then the problem of giving an account of the satisfaction of this condition. It is the problem of explaining the rationality of testimonial cooperation. This is then problematic to the extent that this condition cannot be satisfied; that is, to the extent that we lack reasons—or have a psychological tendency to trust that outstrips our possession of reasons. For the moment, I will leave it open whether, if at all, testimonial cooperation is problematic." (pp. 6-7)

### **3. Differences between the Trust Game and the Testimony Game.**

I think that the differences between the Trust Game and the Testimony Game are too great for Faulkner to be able to draw on the analogy between the two, as he wishes to do. Moreover, the differences between the Trust Game and real-life cases of testimonial exchange, even between strangers, are even greater. I will outline these differences before going on to explore their significance for Faulkner's arguments.

The first crucial difference is that the first player (the investor) makes the first move in the Trust Game, before any other interaction between the two players, and the second player then reacts to the first player's decision. But in the Testimony Game, matters begin with the 'second' player (the speaker) making an assertion; the 'first' player (the audience) then decides whether or not to abandon neutrality of belief, and give some credence to the speaker.

This temporal issue is significant in two ways. The ‘second’ player decides whether to be trustworthy before knowing what the audience will decide to do. This may affect the way people think about what to do: in the lab-based Trust Game, second players are more likely to return cash if the first player has made a relatively generous initial transfer. (This tendency is also illustrated in the Ultimatum Game, in which first players can choose either to keep their stake, or else to transfer some of it to a second player. The second player can either accept what’s transferred, or else reject it, in which case both first and second player lose everything. Second players show a marked tendency to ‘punish’ small initial offers, even at financial cost to themselves.) In testimonial situations, speakers often have to decide whether to be trustworthy even before they know whether their audience will trust them, and cannot react to the audience’s decision.

Moreover the temporal issue means that the audience decides what to do in light of the fact that the speaker has volunteered an assertion. This fact itself is a significant piece of evidence about the speaker, of a kind which is unavailable to first players in the Trust Game.

Ironically, the speaker’s volunteering an assertion makes it more accurate to talk of trust in connection with the Testimony Game than in connection with the so-called Trust Game. Second players in the Trust Game who decide to keep what they’ve been given may perhaps be described as mean, or selfish, or spoilsports, but they are not *untrustworthy*. After all, they make no prior agreement to return any part of the cash, and in general there is no obligation to give money to (non-destitute) people who would like you to, even if they have previously given you money. First players in the Trust Game are not genuinely trusting, they are opting to take a risk: handing over money without being asked, then complaining if it is not returned with interest, is otherwise known as loan-sharking (or sub-prime mortgage mis-selling). By contrast, in the Testimony Game the speaker effectively asks for credence by choosing to make an assertion, and it makes sense to think of the audience trusting or distrusting in response.

The second crucial difference is that, in the Trust Game, ‘never take a risk’ is a reasonable strategy for someone acting as first player with a sequence of different partners. This may not be the income-maximising strategy, but on every occasion the first player will

get to keep the £10, which is not too bad. In the Testimony Game, it is much less clear that ‘never give credence’ is a reasonable long-term strategy for audiences: maintaining neutrality about every proposition you cannot check for yourself will leave you with very few beliefs indeed. This fact may alter the ‘pay-off matrix’ for the Testimony Game, if the relative value of neutrality over false belief shrinks as we increase the number of cases. (One might try to model this in the Trust Game by telling the first player that the stake will be reduced by £1 on each iteration.)

The third crucial difference is that in the Trust Game, the currency (cash) is fully meaningful even in the one-shot version, where players interact just once. In the Testimony Game, as Faulkner makes clear, the ‘currency’ offered by the speaker is not truth but trustworthiness. This is because the currency offered needs to be something which the speaker could benefit from retaining: in a given instance, it might suit the speaker to offer the truth (in boasting about an achievement, for example), whereas trustworthiness involves commitment to speak the truth both when this is convenient and when it is inconvenient. This currency of trustworthiness is thus only fully meaningful when we consider a sequence of interactions.

The fourth crucial difference is that in the Trust Game, players do not switch roles. Even when the interaction is repeated a number of times, individual people stick with their roles as first player or as second player. In real-life testimonial situations, we often switch roles between speaker and audience, even within a given pair: this is otherwise known as a conversation.

#### **4. Consequences of these differences**

Whatever we make of these differences, there is no doubt that Faulkner has given us a fresh, fruitful way of thinking about the challenges we face in testimonial exchange. But what are the consequences of these disanalogies for Faulkner’s broader arguments? So far as I can see, the analogy and attendant problem of cooperation are used in two main ways in *Knowledge on Trust*: to undermine the nonreductionist view that we have a default entitlement to accept testimony, and to make plausible Faulkner’s views about social norms of trust and trustworthiness. I will briefly discuss these in turn.

For Faulkner, the situation of the first player in the Trust Game, and by extension the situation of the audience in the Testimony Game indicates that trust (transferring money, giving credence) is unreasonable unless the first player has some positive reason to think that this will pay off (that the second player will back-transfer money, that the speaker will be trustworthy). But a number of the crucial differences shed some doubt on this. The fact that the ‘second player’ begins, by volunteering an assertion, in the Testimony Game, may mean that the very set-up makes it reasonable for the first player to trust. Admittedly, we might think of this as evidence available to the first speaker, which tells in favour of the reductionist view, but the nonreductionist might instead think of this as a feature of the situation which facilitates default entitlement. Moreover the fact that ‘never give credence’ (unlike ‘never transfer cash’) is not even a moderately-good long term strategy might also tell in favour of a default entitlement. And perhaps the nonreductionist might make something of the fact that we often switch, unpredictably, between the roles of speaker and audience, unlike players of the Trust Game.

Finally, though I find Faulkner’s emphasis on our awareness of social norms of trust and trustworthiness very compelling, I suspect that he is drawing additional, unwarranted support for his view from the rather weak analogy between the two games. Faulkner’s discussion of trust is subtle, and he draws out the importance of normative expectations, of the trustee’s recognition that the truster makes herself dependent through her trust, and of the trustee’s being motivated by concern for this dependency. None of this applies to the Trust Game, in which normative expectations are inappropriate, and the second player’s actions are what count, not her motivations; as I argued above the Trust Game is not really about trust. (In Faulkner’s terms, the Trust Game involves at most predictive trust, not affective trust.) It is in fact more plausible that the Testimony Game involves trust of the rich kind which is governed by social norms, and so it is unclear what, if anything, Faulkner has to gain by beginning with the Trust Game.

**Katherine Hawley**

*University of St Andrews*

kjh5@st-and.ac.uk

## TESTIMONY, TRUST, AND SOCIAL NORMS

Peter J. Graham

Paul Faulkner's book revolves around "the problem of cooperation." According to Faulkner, when we communicate with another as to the facts, we face a situation akin to a prisoner's dilemma (2011: 6). In a prisoner's dilemma, our aggregate well-being will be maximized if we both cooperate. However, given the logic of the situation, it looks like the rational thing to do is defect. We're then faced with a problem: how to ensure the cooperative outcome? A "solution" does just that. Faulkner calls the analogous problem the "testimony game." In the testimony game, there is a sender and a receiver. The receiver wants the truth, but the sender wants to influence what the hearer believes. Given the logic of the situation, it doesn't look like it is always rational for the sender to provide what the receiver requires. The receiver is then faced with a "problem."

Faulkner argues that this problem entails that the hearer must have a reason for thinking the speaker's "purpose in utterance is informative" for "warranted uptake" (2011: 133), no matter how knowledgeable and informative the speaker happens to be (2011: 128). The hearer would need a reason for thinking the speaker, at least on this occasion, is one of the "good guys." The "solution" to the "problem" would then consist in the hearer's reasons for uptake.

Faulkner argues for a "trust-based" solution, where the hearer's attitude of affective trust warrants uptake. When the hearer affectively trusts the speaker, Faulkner argues, the hearer's trust both rationalizes the hearer's reliance and moves the speaker to choose the cooperative outcome, and thereby solves the problem. Hence the title, *Knowledge on Trust*.

I will argue that *the hearer's* attitude of affective trust isn't fundamental. A hearer can be warranted, and even acquire so-called testimonial knowledge, without affectively trusting the speaker in Faulkner's sense. What solves the problem isn't *the hearer's* trust; it's *the speaker's* trustworthiness.

### The Problem of Cooperation

Let's begin by saying more about the problem of cooperation. Since the problem concerns the logic of communication, I shall sometimes call it the *problem of communication*. Faulkner puts the problem in terms of rational choice explanations of behavior:

Suppose...a subject's individual beliefs and desires—his preferences—explain...explain how the subject acts...that rational action aims at the satisfaction of the actor's individual preferences. (2011: 4)

Now the “problem” arises because speakers and hearers have cross-purposes:

[S]peakers and audiences have different interests in communication.... Our interest, qua audience, is learning the truth. Engaging in conversations as to the facts is to our advantage as speakers because it is a means of influencing others: through an audience's acceptance of what we say, we can get an audience to think, feel, and act in specific ways. So our interest, qua speaker, is being believed...because we have a more basic interest in influencing others....[T]he *commitment to telling the truth would not be best for the speaker*. The best outcome for a speaker would be to receive an audience's trust and yet have *the liberty to tell the truth or not*. (2011: 5-6, emphasis added)

So audiences want the truth, but speakers don't always want to provide it. And if we act because of our interests, speakers won't always have audience's interests in mind.

...testimony [can fail to be true] because it was not produced with the intention of getting us, as audience, to believe truly. Giving testimony is something we, as speakers, do for reasons, and *our reasons need not put our audiences' informational needs first*. (2011: 132, emphasis added)

There are two typical reactions to this in the literature. On the first, all it shows is that in many but not all cases, and probably not most, a hearer will need additional supporting evidence to defeat an occasion specific reason for thinking the speaker might actually have chosen to deceive or mislead. Or it may show instead, or in addition, that the hearer should be able to reliably tell when someone is apt to lie or mislead. On the second, this point shows that in each and every case of testimony there is a standing defeater that the hearer must overcome with positive reasons for thinking the speaker is

not likely to lie or mislead. On the first, understanding the speaker is enough for *prima facie, pro tanto* warrant, warrant that sometimes needs supplementation for *on balance* warrant. On the second, understanding the speaker is never, as such, sufficient for *prima facie, pro tanto* warrant; the hearer must always possess, in each and every case, a reason for thinking the speaker is sincere to enjoy any warrant at all, *prima facie* or on balance.

Faulkner seems to fall within the latter camp. He thinks the problem shows that:

....it is not reasonable to trust without a *supporting reason* that rationalizes trust. *The acceptance of testimony must be backed by reasons if it is to be reasonable* [and so knowledge]. (emphasis in original 2011: 6-7; 10, 20-1,114, 118-19)

In short, since communication is like a prisoner's dilemma, the hearer needs a reason for thinking or presuming that the speaker has chosen the cooperative, helpful outcome. Like the prisoner's dilemma where each party knows the other has a reason to defect, the hearer knows in the "testimony game" that the speaker may very well fail to choose the cooperative outcome, for the speaker doesn't, qua speaker, prefer being helpful (even though, as a matter of fact, he may say what's true). The very logic of communication thereby creates a standing defeater that the hearer needs to overcome with positive reasons for thinking the speaker shall choose the cooperative outcome. This is why the hearer (the audience) must have a reason for thinking or presuming that the "speaker's purpose in utterance is informative" for "warranted uptake" (2011: 133). Otherwise it's not rational, *even if the speaker has chosen the cooperative outcome*. The hearer can't simply take it for granted, as it were, that the speaker is one of the good guys, but needs positive reasons for thinking he is.

Faulkner thinks the hearer's attitude of affective trust in the speaker provides just such a reason, and thereby solves the problem. But before turning to Faulkner's "trust-based" solution, I will first sketch a paper from Pettit that I have found helpful in understanding Faulkner.

### **Trust-Responsiveness**

In 'The Cunning of Trust' Pettit first characterizes *interactive, trusting reliance*. A relies on B in this way to the extent that:

1. A relies on B to PHI
2. This reliance is manifest to B
3. A expects B to be well disposed and to attach a greater utility to PHI for the fact that it represents a way of proving reliable. (Pettit 1995: 205-6)

We rely on people in this interactive, trusting way all the time. Those who prove reliable are **trust-reliable**.

Pettit distinguishes two forms of trust-reliability: **trust-worthiness** and **trust-responsiveness**. Trust-worthiness implies that the trustee is antecedently disposed to prove reliable; the trust-worthy person has a trust-worthy *character*. Pettit identifies three forms of trust-worthiness: loyalty, virtue, and prudence. Each provides a reason for a trustor to interactively trust a trustee:

Suppose I believe that someone is a loving family member, a loyal friend, a devoted colleague, or whatever....Or suppose I believe that someone is virtuous; say, a god-fearing sort who can be relied upon to follow certain religious norms...Or suppose I believe that someone is a prudent sort who will see the potential long-term rewards of maintaining a certain relationship...that requires her to prove responsive to certain acts of reliance on my part....[By] manifesting the fact of relying on her...I can actually motivate her to perform accordingly. [These beliefs offer grounds] on which I may expect that if I manifest the fact that I am relying on that person to do something that person will be led to attach a greater utility to doing it. (1995: 208)

If I believe you are loyal to me, or if I believe you follow norms prescribing help, or if I believe that you can see the wisdom in helping, then I would have good reasons rationalizing my reliance on you.

Pettit then explains *trust-responsiveness*. He first notes that human beings, among other things, desire the good opinion of others. We value being loved, liked, acknowledged, respected, admired, and so on. He calls these “attitude-dependent” goods for they depend on being the object of someone else’s positive attitude.

Pettit then argues that our desire for the good opinion of others can give a trustor another reason to trust a trustee. Compactly stated, here’s his argument:

- (1) Suppose Andy would like Lindy to PHI.
- (2) Suppose Lindy desires the good opinion of Andy.
- (3) Suppose Lindy's doing PHI when Andy manifests reliance would earn Andy's good opinion.
- (4) Then Lindy has a motive to PHI when Andy manifests reliance.
- (5) Knowing or believing all of this, Andy would then have a reason to manifest reliance on Lindy.

How does the trustor make his reliance manifest so the trustee will be moved to prove-reliable? By also manifesting a *belief* or *presumption* that the trustee is trustworthy, or at least by manifesting a *disposition* to believe that the trustee is trustworthy should she prove trust-reliable. And to believe or presume that another is so trustworthy is to think well of that person, or at least to think well of that person in the event that he or she proves trust-reliable. By manifesting his reliance, Andy communicates that he believes or presumes Lindy "to be truly the sort of person who will not take advantage of someone who puts" himself at her mercy (1995: 214). This belief, presumption or disposition of the trustor then motivates the trustee to prove reliable. When the act of manifesting reliance interacts "this way with the desire of a good opinion, then the act of trust is likely to have an important motivating aspect for the trustee...It is a sort of bootstraps operation, wherein the trustor takes a risk and, by the very fact of taking that risk, shifts the odds in their own favor" (1995: 215-6). If I trust you in this way, then my trust incentivizes you to be one of the good guys. I then have a positive *prima facie* reason warranting my reliance.

### **The Straight Solution**

So suppose with Faulkner that a hearer's acceptance of a speaker's testimony must be "backed by reasons" (2011: 9-11). A common reaction among epistemologists to this is that the hearer must have a set of warranted *beliefs* that provide an argument from the premise that the speaker has asserted that P to the conclusion that the speaker's purpose is informative, and from there to the conclusion that (probably) P. This argument would overcome the standing defeater posed by the logic of the testimony game. Those beliefs

are then the reasons—or constitute the “having” of reasons—that provide the hearer’s warrant for her testimony-based belief (2011: 119).

Pettit’s paper then provides two different (though related) kinds of reasons a hearer might have that warrant uptake, reasons for thinking that the speaker, at least on this occasion, is one of the good guys. I could possess reasons for antecedently thinking you will prove reliable, reasons from loyalty, virtue or prudence. The solution to the problem of communication would then be straightforward. I would have reasons that “solve” the standing defeater posed by the “problem” of communication. I would have a reason that rationalizes my reliance that also engages your antecedently existing motives for proving reliable.

Or the hearer could possess a “trust-responsive” reason *à la* Pettit. Suppose I need information whether P and I think you might have it. I might choose to ask you whether P and even accept your answer in the belief or presumption that by manifesting my reliance on you I will motivate you to tell me the truth. I would believe that you are able to see that I depend on you proving reliable, and I would believe that you see my dependence as a reason to prove reliable, for you desire my good opinion. I would then believe I have tipped your preferences towards the cooperative outcome. I would then have a reason that rationalizes my reliance that provides you with a new motive for proving reliable.

Even though Faulkner says that both kinds of reasoning just sketched would “solve” the “problem,” and even though the latter mechanism of trust-responsiveness will help us understand Faulkner’s “trust-based” solution, Faulkner rejects both kinds for he thinks they are not *the central reason* we “trust” testimony.

[Requiring reasons like these] over-intellectualizes our relationship to testimony. We do not always base uptake on the belief that what is told is true, sometimes we merely trust a speaker for the truth. ... [These reasons miss] a central reason, arguably the central reason, why we trust testimony... An audience’s reason for the uptake of a speaker’s testimony can be no more than that the audience *believes the speaker, or trusts the speaker for the truth.* (2011: 175-6)

### **Affective vs. Predictive Trust**

To get your mind around this, you'll need to understand Faulkner's distinction between two kinds of trust. Following others, Faulkner distinguishes *affective* from *predictive* trust (2011: 144-147). Trust in both cases is a three-part relation: A trusts S to PHI.

A *predictively* trusts S to do PHI if and only if (1) A depends on S to PHI, and (2) A predicts (believes) that S will PHI. The first condition is factive; A must really depend on S to PHI. The second is not. It's the belief (the prediction) that S will PHI (2011: 145). Talk of trust in this sense "sits happily" with the "intellectual" reasons for uptake just canvassed, for "predictive trust is reasonable just when there are grounds for judging a cooperative outcome; and what makes the act of trusting reasonable is these grounds" (2011: 145).

A *affectively* trusts S to do PHI if and only if (1) A depends on S to PHI, and (2) A expects (has the normative expectation) that (1) will motivate S to PHI (where A [normatively] expects it *of S* that S be moved by the reason to PHI given by A's dependence on S) (2011: 146). Faulkner elaborates on (2):

The normative dimension of the expectation...is then that the trusted party *should be* trustworthy. Thus, in trusting S to PHI, A presumes that S *ought to PHI* and, other things being equal, that S will PHI for this reason. (2011: 147-8)

With this distinction in hand, we are on our way to seeing what Faulkner has in mind. I might *believe* that if I manifest my dependence on you for information whether P, then you will tell me the truth, where my belief is a *prediction* based on reasons or evidence. I would then have a positive reason for thinking you are one of the good guys that overcomes the standing defeater. My uptake is warranted provided I've got good reasons or evidence to support my prediction.

On the other hand, I might *presume* that if I manifest my dependence on you for information whether P that you *ought to* tell me the truth. According to Faulkner, this *presumption* constitutes, or is constituted by, affective trust, the *normative expectation* that you *should* tell the truth. Faulkner's main idea is that this expectation rationalizes reliance and thereby warrants uptake; it provides a "de-intellectualized" solution to the problem of communication.

This move, I believe, marks a significant development in the evolution of Faulkner's thinking about testimonial warrant. I recall his earliest papers using the "problem of communication" to justify a "positive reasons" requirement that predictive reasons would readily satisfy; he seemed to endorse a "straight" solution (Faulkner 2000, 2002). Now he's turned to *trust* as a new basis for warranted uptake. Faulkner's main thesis in *Knowledge on Trust* is that this presumption—this normative expectation—warrants uptake and thereby solves the problem of communication. But what exactly is this presumption, and why does it "solve" the problem? We still have a good amount of unpacking to do to understand Faulkner's "trust-based" solution.

### **The Presumption of Trustworthiness**

Faulkner argues that when the audience A believes that speaker S can see that A is relying on S for information whether P, and in addition A affectively trusts S for that information, then A will make a number of *presumptions*. These presumptions do a good deal of the same work explicit beliefs did in the account inspired by Pettit. As Faulkner sees it, the psychology of the trusting audience goes like this:

1. A *believes* that S recognizes his, A's, trusting dependence on S proving informative.
2. A *presumes* that if S recognizes A's trusting dependence, then S will recognize that A normatively expects S to prove informative.
3. A *presumes* that if S recognizes A's expectation that S *should* prove informative, then other things being equal, S *will* prove informative for this reason.
4. So taking the attitude of affective trust involves *presuming* that the trusted *will* prove trustworthy. (2011: 130)

According to Faulkner, these "presumptions" are not, or "need not amount to," beliefs (2011: 154).

The audience A's reason for believing that p, when this is what S tells him, is not the belief that S will prove to be trust-responsive, it is simply the fact that S told him that p (and A trusts S for truth on this matter). S's telling is seen to provide a reason...because in trusting S for the truth, A

accepts certain things about S and the testimonial situation which yield the presumption that S is trustworthy. (2011: 164)

How do these presumptions “solve” the problem of communication? According to Faulkner, the hearer’s *presumption* that the speaker will prove informative *rationalizes* the hearer’s uptake of the speaker testimony in the same way a hearer’s *belief* that the speaker will prove informative rationalizes uptake. These presumptions then do a good deal of the same work explicit beliefs did in Pettit’s model.

The presumption [that the speaker will prove trustworthy] like the belief with the same content, *makes it probable for A* that p is true given that this is what S tells him. So A’s attitude of trust raises the probability of p [for A], given this is what S tells him. So A’s trusting S for the truth...provides A with an epistemic reason to believe. (2011: 154, emphasis added)

Besides *rationalizing* A’s uptake, does the presumption that the speaker should prove reliable also make A’s uptake *objectively more likely to be true*? It does if, in fact, A’s attitude of trust effectively *moves* the speaker to be informative, or moves the speaker to be more likely to be informative. And, as in Pettit’s analogous model, it often does. Faulkner claims, citing Pettit approvingly, that A’s trust gives S “a reason to be trustworthy...acts of trust can create as well as sustain trusting relations” (2011: 156-7). And so Faulkner thinks that A’s trust not only makes it more probable *for A* that S is apt to prove informative in utterance and so rationalizes A’s uptake, it also *motivates* the speaker to prove trustworthy and thereby raises the objective probability that S will prove informative in utterance; A’s trust is often an *effective* reason—a reason that *motivates* the speaker to prove informative, and thereby objectively warrants an audience’s uptake. We now have a “de-intellectualized” version of Pettit’s “intellectualized” trust-responsive reason for manifesting reliance that fits Faulkner’s requirement that uptake be backed by reasons (2011: 57, 160, 167).<sup>1</sup>

---

<sup>1</sup> Unfortunately Faulkner does not explicate presumptions or show why they are not, or need not be, beliefs. All he says is that when we presume these propositions, we are “not explicitly committing to these propositions in [our] reasoning” (2011: 151). Instead he says that presuming these propositions “partly defines how it is that the attitude of affective trust involves seeing things in a certain light...in the positive light of trust” (2011: 152, 154).

But if a hearer must have the concepts involved to have the presumptions, and so be able to think the presumptions in order to have them, exactly why is this account less “intellectualist” than the analogous solution from Pettit? Granted Faulkner’s account isn’t a goodwill account and doesn’t require

All of this, however, is only a part of Faulkner's picture. For ask yourself the following: why should the *hearer's* normative expectation that the speaker *should* prove reliable actually *motivate* the speaker *to be* reliable? Why should affective trust move the speaker to be one of the good guys? In Pettit's paper the analogous question is this: why should the fact that the trustor has manifested reliance on the trustee motivate the trustee to prove reliable? And in Pettit's paper, the answer is that the trustee desires the positive approval of the trustor. But Faulkner doesn't emphasize—and in fact at times de-emphasizes—the speaker's desire for approval from the hearer as a reason for the speaker to prove reliable.

So what is Faulkner's analogous answer to our question? His main answer—as I understand it—is this: the hearer's normative expectation that the speaker should choose the cooperative outcome is just the hearer's *internalization of the social norm* that speakers should prove trustworthy, *and* the fact that the social norm moves the speaker to choose the cooperative outcome for the speaker has internalized the norm as well. The hearer's trust—the hearer's normative expectation, which *rationalizes* uptake—then “engages,” so to speak, the speaker's internalization of the norm, which thereby *motivates* the speaker to choose the informative outcome.

By my lights, we now have another major development in Faulkner's thinking. We've moved from predictive trust to affective trust, and now from affective trust to social norms. To explain this last development, I need say a few words about social norms and their internalization.

### **Internalized Social Norms**

Social norms are an important species of social institution along with conventions, customs, and laws; they are causal structures with explanatory force. Here are some examples. Different populations and subgroups obviously regularly dress in various ways. In business contexts most people dress business professional. In Muslim societies, most women dress head to toe. But in Western societies women usually wear considerably less. The regularities in behavior are there to see. And these regularities are clearly approved. Business people approve of business professional attire; they disapprove of those who fail to dress for success. Muslim men and women clearly

---

“explicit reasoning” through these presumptions, it doesn't seem to me that Faulkner requires fewer psychological capacities for warrant than Pettit would.

prescribe complete coverage for women, whereas Western men and women care considerably less. There are countless social norms governing human life, varying in a number of ways, from group to group. The existence of social norms, like language, is a human universal (Brown 1991). Some are smart—send kids to school—some are silly—men should wear ties—and some are downright stupid—circumcise your daughters. Though universal, they differ widely. Think of norms governing what foods to eat, and when and where to eat them. But though they differ, at a higher level of abstraction human cultures prescribe and proscribe a lot of the same kinds of behavior. Most prohibit killing, assault, and incest. Most promote sharing, reciprocation, and helping.

A number of disciplines have taken up social norms: sociology, psychology, economics, anthropology, etc. Sociologists and anthropologists take up the whole range of social norms. Economics and political science, on the other hand, have mostly taken up norms that resolve free-rider problems, and in particular free-rider problems that are also many-party prisoner's dilemmas. Something that ensures universal—or even fairly general—cooperation is then a “solution” to the many-party dilemma; everyone is better off if nearly everyone cooperates. Certain social norms are then “solutions” to these dilemmas (Pettit 1990).

Abstractly characterized, social norms are (1) regularities in behavior in a group or population (so they are “norms” in the scientist's sense of what usually or “normally” happens), that (2) are prescribed by members of the population, so that they approve of conformity and disapprove of deviance (so they are “norms” in the philosopher's sense of what's prescribed or what ought to occur), and that (3) are regularities in part because they are prescribed (so that they are “normative” in the moral psychologist's sense of motivating behavior) (Pettit 1990, Miller 2001). These are general claims about most people in the population: most follow the norm; most prescribe the norm; and most follow the norm partly because prescribed.

Social norms are experienced as things we *ought* to do, that it would be *wrong* not to do. And so when we prescribe conformity, we believe each of us *ought* to conform, and when someone fails to conform we experience more than mere disapproval, but a kind of *moral* disapproval (Miller 2001: 139; cf. Pettit 1990). Think of the norm forbidding eating pork in certain communities. Sure, eating pork will win

disapproval, but many who conform believe they *shouldn't* eat pork, that eating pork is *wrong*, no matter what other people might think.

In other words, we *internalize* norms. We internalize norms through socialization, from our parents and others. When we internalize a norm, we find it intrinsically motivating; our preferences change. We conform because we think it's the *right* thing to do, because we are *supposed* to do it. We *want* to do it. We may even deeply value compliance. Many internalized norms even come to be partly “constitutive of the selfhood or identity of individual adherents” (Miller 2001: 139). Internalization then leads to compliance as an ultimate end, and not just as a means to avoid punishment or societal disapproval (Bowles & Gintis 2003: 13-14, 2011: 169). When internalized, reward and punishment may drop away, or only play a sustaining role. When internalized, I conformed to the norm *because it's the right thing to do*, because I positively value compliance, not (normally or just) because of the consequences of my actions or because of my other aims or desires. Internalized norms are then experienced as categorical, as what must be done, and not simply as what we should do given other aims or desires (Sripada & Stich 2006). Given that I've internalized a norm, I expect it of myself and others.

Internalized norms are intimately connected with the social emotions: guilt, shame, embarrassment, love, envy, pride and resentment. These emotions make a huge part of the proximate psychological mechanisms driving positive and negative evaluations of compliance, evaluations that motivate compliance; strong social emotions are a central proximate psychological mechanism ensuring conformity (Frank 1987; Ekman 1992). Even thinking about failure may evoke strong feelings of embarrassment, anxiety, guilt or shame (Gintis 2003). You might even feel sick at the very thought of breaking a norm. If I fall short of my normative expectation of myself, I will feel guilty or ashamed. If you fall short, I will feel contempt or resentment. “Social norms have a grip on the mind that is due to the strong emotions they can trigger” (Elster 1999: 100).

### **The Ultimatum Game**

Faulkner thinks we've internalized social norm *tell the truth informatively*. And he thinks that fact somehow “solves” the problem of communication, for it rationalizes uptake and motivates trustworthiness. To explain how this works, I'll first explain how

internalized norms drive behavior in another game studied by rational choice theorists, economists, anthropologists and evolutionary psychologists: the ultimatum game.

In the ultimatum game, there are two players, the proposer and the responder. The proposer is given some money, e.g. \$100. The proposer then must propose a split of the money between the proposer and the responder, anywhere between \$1 for the responder and \$99 for the proposer and \$99 for the responder and \$1 for the proposer. The responder's job is to accept or refuse the split. If accepted, both parties receive the amount in the proposal. If refused, no one gets anything. And so both parties are better off if the responder accepts the split. These games are often played anonymously, where both parties know they won't play another round with the same person, and so tit-for-tat, among other strategies for playing repeated games, isn't an option.

According to a simple rational choice prediction where people rationally act on their preferences—especially their economic ones—the proposer should propose \$1 for the responder and \$99 for the proposer, for the proposer will know that the responder will accept the split, for the responder will reason that \$1 is better than nothing, and so it is better to accept such a split than refuse it.

Surprisingly, however, across a very wide-variety of human cultures, that is not what happens. Instead the proposer tends to propose something more like \$40 for the responder and \$60 for the proposer (the split on average varies from culture to culture), and when the proposer proposes a much smaller split for the responder, the responder tends to refuse them, which in part would explain why bigger splits are usually proposed, for the proposer usually knows which splits are apt to be refused, and which are apt to be accepted.

A now standard explanation of this behavior—which seemingly violates “rational choice axioms” of human behavior—is that the participants—and so humans generally—have internalized social norms of fair-divisions of goods (Henrich et al, 2004). Proposers know that they are supposed to propose a fair split, and responders know that too, and furthermore responders get upset and enforce the norm by punishing the proposer by rejecting obviously unfair splits (and thereby also harm themselves, for they lose out on the money offered in the split).

Describe the game to your colleagues, students, or loved ones, and you'll elicit similar reactions. Sure, some of them might take a low offer when playing responder,

but many are bound to feel punitive attitudes—to get upset—at “unfair” offers. They “know” the offers are *supposed* to be fair, that the proposer *should* offer a more even division of the goods. Because of internalized norms, we normatively expect fair divisions, and so fair divisions reliably occur.

If such “fairness” norms have been internalized, they explain why people behave as they do when playing the ultimatum game. We then have an example of where internalized social norms explain behavior that otherwise would have seemed very puzzling, at least from the point of view of rational choice theory. We have an example of behavior that goes against what rational choice theory would have predicted in the first place. Participants “see” what they are “supposed” to do in such a situation in terms of internalized norms, and those internalized norms then explain, in large part, what they do.

Now when playing the ultimatum game it’s often the proposer’s recognition that the responder expects a fair division that at least partly moves the proposer to offer a fair split. But just as often the proposer herself thinks she should offer a fair split, for she has internalized the norm, and believes anything less would be wrong. She might even, and often does, propose a fair split in games where she knows the game is anonymous and won’t be repeated. She “intrinsically values” fair divisions of the goods.

Social norms also enter explanations of actual behavior in prisoner’s dilemmas. Given a rational choice model, you’d expect everyone to defect, especially in non-repeated, anonymous games. But in fact a substantial number of people, even in non-repeated, anonymous games, choose to cooperate, and they often do so because they believe it’s the right thing to do.

### **The “Trust-Based” Solution**

Now recall our question for Faulkner: What is it about the *hearer’s* normative expectation that the speaker should prove reliable that motivates *the speaker* to be reliable? And recall our answer: the hearer’s normative expectation is the hearer’s *internalization of the social norm* that speakers should prove trustworthy, and if the speaker has internalized that norm, then the speaker’s recognition of the hearer’s reliance shall move the speaker to prove reliable. The *hearer’s* internalization of the norm then “engages” the *speaker’s* internalization of the norm, which motivates the

speaker to choose the informative outcome. The speaker is then one of the good guys, for he prefers helping. We're now in a position to say what all of that means.

Faulkner holds that when we play the testimony game we follow the internalized social norm *tell the truth informatively*.

We expect interlocutors to live up to [a standard] when...having a certain type of conversation. On this standard if another depends on you for information, then you should try to say what is true informatively. (2011: 181)

Once internalized the norm shapes our moral psychology. It shapes how we "see" the situation, our behavior, and our reasoning. Faulkner says this is a...

...quasi-perceptual matter because where the norms of trust are internalized, the subject's perception ...[of]...the Testimony Game will be structured by the prescriptions of the norms. This situation will be *seen in a certain light*. Thus, *if S can see that A depends on her telling the truth as to whether P, this will be seen by S as a reason to do so...*(2011: 185, emphasis added).

Speakers who have internalized these norms—and so intrinsically value compliance—will then often enough choose the informative outcome when they see that audiences need information; they will be "motivated to conform" because they have "internalized the norm" and so "intrinsically value" compliance (2011: 186). Speakers are then committed to telling the truth; their reasons for utterance will then often put audience's informational needs first (cp. Faulkner 2011: 5-6, 132). And so when playing the testimony game, audiences that manifest reliance will receive the cooperative outcome, for speakers will be motivated by manifest need, for they intrinsically value proving informative.

We can then explain the matter of fact reliability of our institution of testimony:

Our ability...to see testimony in the light of trust comes down the fact that we have internalized social norms of trust. ...It is these...*social norms that shape the motivations we have, as speakers, in giving testimony, and which determine that testimony is the reliable source of knowledge that it is*. (2011: 172, 51, emphasis added)

By Faulkner's lights, we have the materials in place for a "genuinely philosophical theory of testimony," for that requires an "explanation as to why testimony can be presumed to be reliable and reliable in the way that it is" (2011: 51).

This explanation then "solves" the "problem" of cooperation:

[T]he *solution* to the problem [of cooperation in the case of communication comes from our] social institution of testimony defined by the existence of social norms. (2011: 169, emphasis added)

We can now see why uptake is warranted. It's warranted because *speakers* have among their preferences the pro-social, cooperative, helping preference to prove informative in communication.

This not only solves the problem, it "dissolves" the problem. Recall how the problem was set-up. When playing the testimony game, hearers want true information. But speakers supposedly would prefer the liberty to speak as they please; they are not committed to telling the truth; their reasons for utterance do not include the audience's informational needs. It then looks like the hearer needs some reason for thinking the speaker has chosen the informative outcome before it is rational for the hearer to rely on the speaker's testimony; the hearer needs some reason for thinking this speaker is one of the good guys. But if we've internalized social norms of trustworthiness, there's no problem, for it's not true that speakers are just as apt to prove uninformative as informative. It's not true that they're out simply to influence our beliefs. Rather they are out to inform, to prove helpful; speakers *intrinsically value* informative outcomes; they are committed to informative outcomes. We're then not in a world where, from the point of view of the logic of communication, the world might as well be equally divided between the good guys and the bad guys where we need evidence telling them apart; rather we're in a world of good guys.

...the trust-based solution...*dissolves* the problem of cooperation through showing how it rests on a restricted conception of what reasons we can have for ...being trustworthy....*it shows how trust need not be conceived of as problematic...*[when] norms of trust and trustworthiness are internalized. Then the Testimony Game would simply be perceived by interlocutors as a situation wherein each had a reason to trust and be trustworthy....*the trust-based solution...is not a strategy...for removing a background of distrust...*Rather, the claim that we intrinsically value

trust and trustworthiness is part of a philosophical explanation as to why such distrust is not pervasive. (2011: 199, emphasis added)

Like the ultimatum game, when we play the testimony game we don't simply have the kinds of motivations rational choice theory would predict. Behavior that had seemed problematic in the first place isn't so problematic after all. The existence of internalized social norms "solves" the "problem" of communication.

We should now be able to appreciate exactly why, on Faulkner's view, the hearer's attitude of affective trust warrants uptake:

...ultimately what determines that the attitude of trust provides an epistemic reason is *the existence of the causal structures* [the social norms prescribing truth telling and trusting] that *ensure* the attitude of trust can be potential evidence [in the sense that given the attitude, it is objectively more likely that what the speaker says is true]...Thus, ultimately, the attitude of trust provides an epistemic reason because there are norms of conversational trust, shaping the nature of the reasons we have, for utterance and for belief, in conversations as to the facts. (2011: 159, emphasis added)

The "ultimate," fundamental basis for the solution to the problem of communication—and so the ultimate, fundamental basis for warranted uptake—rests on the internalization of social norms.

There is much to admire in Faulkner's "dissolution." As Faulkner is aware, and as the discussion of the ultimatum game conveys, the appeal to social norms is one of the mechanisms social scientists now appeal to when explaining why humans cooperate—why they help one another—to the extent that they do, especially when a prediction based on narrow self-interest would have predicted something else instead. Faulkner's application of the apparatus of social norms to the case of informative cooperation is then an instance of a more general strategy for explaining why humans are as "pro-social" in their behavior as they are. I've pursued the very same strategy in some of my own work on testimonial warrant (see Graham, forthcoming, in preparation; cp. Graham 2010).

### **Pulling the Rug Out**

I shall make one major criticism. Given the way I understand it, to dissolve the problem *speakers* must have the hearer's interests in getting the truth as one of their interests. If *speakers* don't have that interest, then we have a problem. If they do have that interest, then we don't. If *speakers* have internalized the social norm *tell the truth informatively*, then they intrinsically value the informative, cooperative outcome, and then have providing truth to others as one of their interests; they then have "pro-social, cooperative" preferences. So suppose, with Faulkner, that we have internalized that norm. Then when playing the role of speaker in a testimony game, we'll have among our preferences telling the truth informatively. Then when the hearer asks whether P, we will provide the truth informatively, insofar as we are able. We've thereby "dissolved" the problem.

I have identified two major roles the hearer's affective trust—the hearer's normative expectation that the speaker should prove reliable—plays in Faulkner's account. The first is *rationalizing*. The hearer's belief that the speaker can see the hearer's reliance and can recognize the hearer's expectation, plus the nested presumptions that the speaker will then prove reliable, all rationalize or render reasonable for the hearer the hearer's uptake of the speaker's testimony; it overcomes a standing defeater. The second is motivational. Like Pettit's mechanism of trust-responsiveness, the hearer's normative expectation of the speaker can play a motivating role; it can motivate the speaker to prove reliable. It then renders objectively probable the hearer's uptake.

But if we have *dissolved* the problem through the speakers internalizing the norm *tell the truth informatively*, why must the hearer also internalize the norm? Why is the hearer's internalization of the norm fundamental in explaining warranted uptake?

Consider the rationalizing role. The demand for reasons in each and every case was motivated by thoughts about speaker preferences: speakers act on their self-interested preference to influence the hearer's beliefs, not on any other-regarding preference. But now we know that's not true, for we know speakers have other-regarding, pro-social preferences as well, for they internalize the norm *tell the truth informatively*, and thereby intrinsically value proving informative. They are committed to telling the truth and have the audience's needs among their reasons. So the basis for

the “problem” of communication has been removed, thereby “dissolving” the problem. Sure, we may need reasons to rebut occasion specific reasons for thinking this or that speaker might actually lie or mislead, but we don’t need a reason in each and every case to rebut a standing defeater, for the very reason for thinking there was such a defeater falsely assumed speakers only care about influencing what hearers believe and never care about what hearers want when playing the testimony game. Once speakers internalize the norm, we’ve removed a basis for “distrust.” Once dissolved, Faulkner’s motivation for requiring reasons for warranted uptake goes by the board.

Consider next the motivating role. If speakers have internalized the norm, then they are already motivated to prove informative in communication. They don’t need, at least not in each and every case, further motivation from the hearer’s normative expectation that they should prove reliable to be reliable. As Faulkner says, it’s the fact that speakers have internalized the norm that explains the reliability of testimony. If they haven’t, then we’re in trouble. But if they have, hearers needn’t worry, at least not in each and every case. And so they needn’t normatively expect speakers to be reliable in every case to motivate them to be reliable. Once we’ve shifted from something like Pettit’s mechanism—where something in the trustor explains the trustee’s motivation—to internalized social norms as the mechanism motivating trustworthy behavior, then the need for something like Pettit’s mechanism in each and every case drops away for warranted uptake, uptake not backed by other reasons either. So when Faulkner turns to internalized norms to explain why the hearer’s attitude motivates the speaker, he’s effectively shifted away from a reason that *the hearer* must possess to explain warranted uptake to a motive *the speaker* must possess to explain warranted uptake.

In sum, Faulkner has shifted from the hearer’s attitude as fundamental to the speaker’s motivation as fundamental; supposedly the hearer’s attitude did a lot of the work, but now we can see that the speaker’s motive does all the heavy lifting. This shift dissolves the basis for both the rationalizing and motivating roles of the hearer’s normative expectation. As far as I can tell, Faulkner has pulled the rug out from beneath his feet.

### **The Problem of Parental Care**

We even know such a rationalizing and motivating reason based in the recipient isn't required to motivate pro-social, helpful behavior.

Consider parental investment in offspring. As any parent knows, parents invest an enormous amount of resources in the care of their children, even when the children are unlikely to repay any of it. From the point of view of narrow self-interest, parental investment in their offspring can seem puzzling. So why do they do it?

Well, again as every parent knows, parents deeply care about the well-being of their children. Parents have "internalized" the well-being of their offspring; they have very strong preferences in favor of helping their children. That in part explains why parents invest so much.

Now what do children have to do to get their parents to invest so much care? Nothing really. They just have to need the help. They certainly don't have to manifest a belief, presumption, or disposition that if their parents care for them, they will in turn pay them back, or confer some other benefit. They certainly don't have to manifest the normative expectation that their parents *should* invest so much in them, and they certainly don't need a reason for thinking that their parents will help them. Newborns, infants, toddlers and very young children are not even psychologically capable of possessing these mental states and attitudes.

And so here we have a case of helping behavior explained by helping preferences. The children don't need to normatively expect anything from their parents. And so the "solution" to the "problem of parental investment" does not require some "rationalizing" and "motivating" attitude that "warrants" children's "uptake." The solution consists in parents caring for their children.

And so when it comes to parents talking to their children, the children obviously do not need to motivate their parents, or rationalize their trust, via a normative expectation that their parents should prove reliable. Parental testimony is an existence proof of the possibility of testimonial warrant and knowledge in virtue of the speaker's motivations without a corresponding normative expectation on the part of the recipient. Parents regularly and reliably provide true information to their children because they care about the well-being of their children. Relatives, teachers, and other caregivers too. Even complete strangers often provide true information to children. But children—

especially very young children—don't possess reasons that either motivate or rationalize uptake. At best they sometimes have grounds for suspending judgment (Graham 2010). They certainly don't possess the normatively loaded attitudes Faulkner has in mind. Nonetheless, children learn (come to know) a great deal by believing what their parents, relatives, teachers, caregivers and even strangers tell them. Children thus don't need to internalize social norms of trust and truth telling to rationalize their reliance or motivate other people to tell them the truth. If parental concern can explain childhood knowledge through testimony without a corresponding normative expectation on the child's part, why can't a speaker's concern explain adult knowledge through testimony without a corresponding normative expectation on the hearer's part?

On the last two pages of his book Faulkner addresses the problem posed by children. In total, he devotes less than a page to the problem. And what he says strikes me as unsatisfactory.

Firstly, he says that children are not entirely unsophisticated, in that in communication they are sensitive, to at least some degree, to the speaker's communicative intention (which, on standard accounts of linguistic comprehension, is required to understand speaker meaning) (2011: 203, n. 3). But sensitivity to communicative intentions is not the same as affectively trusting a speaker for the truth; at best it's a prerequisite for that. He does not say children are capable of affectively trusting speakers in his sense. He does not say they are capable of believing that the speaker can recognize their dependence. Nor does he say they are capable of presuming that the speaker *should* prove trustworthy in communication. And so he does not say that children possess the reason he is so at pains to locate as *the central basis* for our rational reliance on testimony. Furthermore, he does not say that children possess either kind of predictive reason that Pettit isolates.

Secondly, even if children are sensitive to the speaker's communicative intentions, Faulkner says nothing as to why that should warrant uptake. Indeed, if it did, Faulkner's own account would be otiose, as every single hearer must recognize communicative intentions in the first place to understand the speaker, and then every single hearer would at least enjoy a *prima facie, pro tanto* warrant via understanding. Furthermore, Faulkner seems to repeatedly reject this position, as sensitivity to communicative intentions is already present in the testimony game. Mere sensitivity to

communicative intentions, Faulkner thinks, isn't enough to warrant uptake. Faulkner barely mentions what children can do, and he says nothing about why it would be adequate. And the one thing he does say they can do he rejects as inadequate.

Thirdly, he then makes a familiar move when pressed by such cases: deny the data. If children are not sophisticated enough to have the reasons Faulkner requires, then, Faulkner claims, their testimony-based beliefs really aren't warranted, and they don't really acquire knowledge from testimony, no matter how informative their parents and caregivers happen to be. But where is it written that children cannot acquire knowledge from their parents and caregivers unless they've matured enough to have the kinds of attitudes Faulkner requires? "There is a vague domain here," he writes, "and the objection should not illicitly exploit it" (2011: 203). The vague domain is the point at which they go from unsophisticated in the relevant domain to sophisticated. The vague point instead, Faulkner would have us believe, is when they go from ignorant recipient of highly reliable information to knowing recipients of that very same information. The objection does not illicitly exploit the first. Rather, Faulkner (at least it seems to me) illicitly conflates the former with the latter.

The case of childhood testimony is then an existence proof that the speaker's motivations are often sufficient for testimonial knowledge and warrant, that a hearer does not always need positive supporting reasons to warrant uptake. Pro-social, helping preferences are often sufficient to explain warranted uptake; reasons for uptake are not required in each and every case.

### **Why not Reliabilism?**

My objection to Faulkner is that I don't see why we, qua hearers, must internalize the norm (though clearly, in general, we have). And so even though obviously present, I don't see that Faulkner has explained why it's fundamental, when it seems, on his account, that it's the speaker's internalization of the norm that is both necessary and sufficient. It can't be that, qua hearers, we need a reason to think that the speaker has chosen the informative outcome, for once the norms are in place, the problem of communication has been dissolved, and so internalization qua hearer isn't required to overcome a standing defeater; it is simply not true that speakers are on the fence.

On the other hand, I don't deny that we, qua hearers, have internalized the norm, and I don't deny that, given internalization, it makes sense "from the inside" for hearers to rely on speakers to prove informative, and even to normatively expect speakers to prove reliable. What I deny is that Faulkner has established the requirement that, in every case, the hearer must have internalized such a norm, and so that the hearer must, in every case, presume that the speaker ought to prove reliable and so will be reliable, in order for the hearer to enjoy a *prima facie, pro tanto* warrant to rely on the speaker.

Turning to the issue of motivation, it can't be that, qua hearers, we need to internalize the norm in every case to motivate others, qua speakers, for, qua speakers, we will choose the cooperative outcome, provided we have internalized the norm; the norm in speakers, as it were, doesn't get turned on when and only when they recognize the internalization in hearers; it gets turned on when they recognize that others need information. A trust-responsive mechanism isn't required, for speakers are antecedently motivated to prove informative.

For instance, I might think you need information whether P and provide it. You might, however, not need it, or not care whether I provide it. You might even suffer from brain damage that left you intelligent but affectless, so that you no longer experience the emotions characteristic of internalized norms. As a result you could care less whether anyone provides true information informatively. I might even know that you don't care. Even so, given I've internalized the norm, I still tell you the truth informatively; it's the right thing to do.

So once we've internalized the norm *tell the truth informatively, qua speakers*, why must we internalize the norm, *qua hearers*, to either motivate the speaker or rationalize our reliance, or both? It's not required to motivate speaker compliance, and it's not required to rationalize reliance, even if it does. And so why not choose the reliabilist, externalist account where hearers are *prima facie, pro tanto* warranted in believing what they are told because testimony is reliable, reliable because, qua speakers, we've internalized the norm *tell the truth informatively*?

Faulkner's own thinking has evolved from a straight solution to something modeled on Pettit's reasoning to the social norms account. As I've argued, once he's moved to the social norms account, he's pulled the rug out under his feet; he no longer

has a basis for insisting that the hearer must have, in each and every case, a reason warranting uptake. Speaker reliability, grounded in pro-social preferences, is enough.<sup>2</sup>

**Peter J. Graham**

*University of California, Riverside*

## References

- Bowles, S. & H. Gintis (2003) 'The Origins of Human Cooperation' P. Hammerstein, ed. *The Genetic and Cultural Origins of Cooperation* (The MIT Press).
- Brown, D. (1991) *Human Universals* (Philadelphia: Temple University Press).
- Ekman, P. (1992) *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage* (New York: W.W. Norton).
- Elster, J. (1999) *Alchemies of the Mind* (Cambridge University Press).
- Faulkner, P. (2000) 'The Social Character of Testimonial Knowledge' *The Journal of Philosophy* 97.
- Faulkner, P. (2002) 'The Rationality of Our Reliance on Testimony' *Synthese* 131.
- Faulkner, P. (2011) *Knowledge on Trust* (Oxford University Press).
- Frank, R. (1987) *Passions within Reason* (New York: W.W. Norton).
- Gintis, H. (2003) 'Solving the Puzzle of Human Prosociality' *Rationality and Society* 15: 155-187.
- Graham, P. J. (2010) 'Testimonial Entitlement and the Function of Comprehension' A. Haddock, A. Millar, D. Pritchards, eds., *Social Epistemology* (Oxford University Press).

---

<sup>2</sup> I have benefited from comments from Coleen Macnamara and Zachary Bachman.

- Graham, P.J. (forthcoming) 'Epistemic Normativity and Social Norms' D. Henderson & J. Greco, eds, *Epistemic Evaluation* (Oxford University Press).
- Graham, P.J. (in preparation) 'The Reliability of Testimony and Social Norms'.
- Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr & H. Gintis (2004) *Foundations of Human Sociality* (Oxford University Press).
- Miller, S. (2001) *Social Action* (Cambridge University Press).
- Pettit, P. (1990) 'Virtus Normativa: Rational Choice Perspectives' *Ethics* 100: 725-755.
- Pettit, P. (1995) 'The Cunning of Trust' *Philosophy and Public Affairs* 24: 202-225.
- Sripada, C. & S. Stich (2006) 'A Framework for the Psychology of Norms' P. Carruthers, S. Laurence, & S. Stich, eds., *The Innate Mind*, Volume 2 (Oxford University Press).

## REPLIES

**Paul Faulkner**

In the following sections I try to offer some reply to the contributions made to this *Abstracta* edition discussing *Knowledge on Trust*. For this opportunity of conversation I am extremely grateful and would like to thank both the contributors and the editors of *Abstracta*.

### **Reply to Longworth**

I'd like to concentrate my reply to Longworth along three axes: my account of how we satisfy the demands of doxastic responsibility since this is the only place where I think Longworth does not get me exactly right; the epistemic externalism of my account since Longworth thinks that this is problematic; and the presumption I make about the nature of practical reason since this is the nub on which everything turns, as Longworth is correct to observe.

First, doxastic responsibility requires that the uptake of a piece of testimony be based on reasons. This condition, I think, is established by the problem of cooperation; it is established by the fact that there are a multitude of potential explanations of any given bit of testimony. As Longworth observes, I think that this tells against the non-reductive theories of McDowell and Burge. The reasons required for responsible uptake must be available *in advance* of that uptake, contrary to McDowell. And the origin of this demand for reasons implies that there can be no *pro tanto* entitlement, contrary to Burge. Given that I reject these options, Longworth observes that an alternative account of how uptake satisfies the demands of doxastic responsibility is required. And then notes that the account I offer suffers the same problem as McDowell's account; that is, even if trust can serve as evidence of a speaker's trustworthiness "by conditioning them to be trustworthy", it is "not clear how that evidence could ground the *initiation* of that trust." (n.10, p.24).

Questions of doxastic responsibility, I think, Longworth is right to observe, are to be settled by appeal only to internalist reasons; or reasons that are reflectively accessible. Doxastic responsibility tracks how a subject thinks about a situation, which in this case is the potential uptake of a piece of testimony. As such, it is not an audience's attitude of trust *qua* piece of evidence, if indeed it is that, which satisfies the demands of doxastic responsibility. Rather, it is the audience's attitude of trust itself, and specifically the presumptions that this attitude carries with it – how the audience thinks about the trust situation – that satisfies these demands. It does so because it identifies one of the multitude of potential explanations of a bit of testimony as the actual explanation; and because there is a potential explanatory connection between this act of selection and the speaker's trustworthiness. In this respect, the attitude of trust shares a property with something that is a bit of evidence for a speaker's trustworthiness. And it is its sharing this property that makes it fit for serving an epistemic role, which is rendering uptake doxastically responsible. So while it is true that an actual explanatory connection between trust and trustworthiness could not ground the *initiation* of trust, what grounds this initiation is not this connection but how the subject thinks about the trust situation in adopting the attitude of trust (with the presumptions it expresses).

It is worth elaborating this point somewhat. A full account of our reasons for uptake should, I think, distinguish between our situation, and the genealogically basic case characterized in the problem of cooperation. In our case, the demands of doxastic responsibility can be satisfied by all the empirical reasons that we have for thinking that a bit of testimony is true. These reasons are numerous. However we possess these reasons only insofar as uniformities exist within the testimonial domain, and this requires uniformities in intention. Our situation is one where there are *norms of trust* that determine these uniformities. These norms also determine a potential explanatory connection, by way of common cause, between trust and trustworthiness. (This explanation ordinarily follows the logic of the norm but it can also be mediated by the desire to avoid the punitive attitudes that are associated with norm transgression.) And these norms determine a speaker's expectation that an audience's reason for uptake be trust – or faith as Longworth calls it – and not empirical judgement. So our situation is one where we live in two 'worlds': we

have many empirical reasons to believe but are often called to believe on the basis of trust or faith; and where the tension between these reasons is also felt when a speaker asks to be given the benefit of the doubt, or when trust can otherwise take one out on a limb. The situation in the genealogically basic case is much bleaker. There is not yet the normative structure to ground either the presumptions of trust or the uniformities needed for the possession of empirical reasons. The problem of cooperation thereby confronted is then only resolved through an intrinsic valuation of trust and trustworthiness, where this evaluation is achieved in our case through our thinking about trust in the manner encoded in the norms of trust. This is the normative structure that then provides for both trust and empirical reasons. So the focus on trust is due to its role in the provision of reason, in setting up our world where empirical reasons are then bountiful and trust often, but not always, falls into the background.

Second, the idea that knowledge can be got on trust forces one to adopt an epistemic externalism about knowledge. Longworth argues for this consequence as follows. Call the case where a speaker knows what she says ‘the good case’, and the case which differs only in that the speaker does not so know, ‘the bad case’. The internalist reasons that support uptake in the good case – be they empirical reasons or trust – cannot outstrip the reasons available in the matching bad case. So if one can get to know something on the basis of testimony in the good case, that knowledge must be accounted for in externalist terms. This argument is good, I think; but the case for an externalist account of testimonial knowledge can be put more directly. Suppose, on the basis of testimony, I get to believe that spider silk is five times stronger than steel.<sup>1</sup> For this to be an item of knowledge, someone has to have done the science needed to establish it. However, the reason I possess for believing that spider silk is five times stronger than steel might be no more than that I trust some source. But if this is so, my belief is not based on my having done the science, so someone other than me must have done this science if I am to know this claim to be true. So if I do possess this knowledge, at least part of the explanation of my knowing this must refer to this science. This reference makes the account of my (testimonial) knowledge externalist.

---

<sup>1</sup> <http://www.guardian.co.uk/science/video/2012/jul/06/unravel-secret-spider-silk-video> . Accessed 9/7/12.

What is problematic here, according to Longworth, is that this externalism is at odds with the claim that questions of doxastic responsibility must be settled only by reference to internalist reasons.

It would be natural to expect facts about what one knows or doesn't know to play a role in determining whether or not one meets the demands of doxastic responsibility. ... However, with respect to the cases of knowledge that are of greatest interest to Faulkner—the normal cases in which one acquires knowledge from another—the expectation is not met. Some of the reasons that play an essential role in determining whether or not one knows are externalist reasons. And those externalist reasons are precluded from figuring in the determination of whether or not one's beliefs meet required standards of doxastic responsibility. On Faulkner's account, wherever one knows just on the basis of (responsibly) accepting what one is told, it is one's belief only, and not one's knowledge, that figures in determining whether one is, in further cases, being responsible in believing as one does. (p.25)

The solution, Longworth thinks, is to assert that the reasons made available by the good case are different to those made available by the bad case. However, this solution is McDowell's and adopting it means that the phenomenological objection I press on McDowell – against his account of our reasons for testimonial uptake – no longer stands.

The solution Longworth proposes, following McDowell, shows that epistemological externalism is not forced upon us by the suggested argument; that is, by the fact that we can imagine cases that are good and bad in the described way. However, epistemological externalism is independently plausible because of cases like the one stated, whereas the strategy Longworth proposes for avoiding externalism becomes less plausible when applied to such cases. Take the good case to be the actual case – wherein I get to know that spider's silk is five times stronger than steel – and the bad case to be one where the background scientific calculations and experiments were done badly (the result is still announced to be the same but the science does not establish it). The proposal is that the reasons that are reflectively accessible in the good case are different to those that are reflectively accessible in the bad case even when the only difference between the cases is that the scientific calculations were done badly in the latter case. Since this difference is not one that the

audience is cognisant of at any level – if he were, the bad case would not be matching – any sense of ‘reflectively accessible’ in play is implausibly etiolated. Moreover, and importantly for Longworth’s objection, it is impossible to see how this difference in what could be reflectively accessed could make a difference with respect to questions of doxastic responsibility such that the audience is rendered responsible when further uptake is based on the good case belief but not the bad case one. But if this is so, it cannot be a unique criticism of externalist accounts – and in particular my externalist account – that matters of external difference might equally fail to make such a difference. The problem is general to any account that allows testimonial knowledge of scientific claims (or claims that require a specific kind of warrant, which may not be possessed by an audience).

Third, the argument I give against Burge requires a specific conception of the nature of practical reason. As Longworth observes, it is the connection Burge presumes between belief and speech that I take issue with, and specifically the presumption of sincerity. One might trust a speaker to be sincere, and have reasons for thinking that the speaker is so, but there is no entitlement to presume sincerity established merely by the rationality of speech. This is because speaking is a practical activity and so responsive to the speaker’s interests, which need not be the same as the audience’s interests. Moreover, with respect to conversations as to the facts, the primary concern of any epistemology of testimony, if there is any presumption, it is that there will be such a divergence of interest. This is the root of the problem of cooperation, and it is this problem that Burge’s argument does not adequately address. Longworth’s acute observation is then that this problem presupposes a conception of practical reason – servicing an individual’s ends – to which one might object. In particular, there is an alternative conception according to which the primary task of practical reason “is the selection and attainment of ends that are *good*” (p.28). So its first task is working out what is moral and its second task is then servicing this end. If this account is adopted *and* it is proposed that there is “a *pro tanto* moral requirement to the effect that one should do one’s best to service an audience’s epistemic ends” (p.29), then Burge’s argument goes through because “[w]e would be entitled to faith, not only in others’ epistemic capacities, but also in the goodness of their wills.” (p.30)

This much, I think, is true: in trusting others, we do rely on the goodness of their wills; and in thinking about the trust situation as the trusting and trustworthy person do we take it that there is *pro tanto* moral requirement to the effect that one should do one's best to service an audience's epistemic ends. This requirement is effectively a particular expression of what I have termed the *norm of trustworthiness*. In thinking in terms of this norm, we think in a way that suggests the truth of the stronger conception of practical reason that Longworth, following Aristotle and Kant, proposes. However, the issue is whether this conception of practical reason is true, or merely a true description of how we think. The problem, I think, is that when we imagine the genealogically basic case what we imagine is a state of affairs where this conception of practical reason does not hold. The result is that we confront the problem of cooperation and what then matters is securing the motivations needed to solve this problem and so make society a possibility. Doing this involves finding some way of intrinsically valuing trusting and trustworthy behaviour. The way that we have secured these motivations is via a certain way of thinking about trust and trustworthiness, where this way of thinking is encoded in our norms of trust and articulated by the stated moral requirement. So *given* our set of thick ethical concepts, there can be the presumption of others' goodwill that suggests that the stronger conception of practical reason is true of us. But here my sympathies, as elsewhere, lie with Bernard Williams: we could have had a different set of thick ethical concepts. What matters is securing the motivations necessary for resolving the problem of cooperation, not how these motivations are secured. It then seems plausible that not all ways of securing these motivations would result in anything like our norm of trustworthiness or the stated moral requirement. But if this is the case, it would seem that the stronger conception of practical reason is no more than a truth about us, and the minimal more egoistic conception presumed in setting up the problem of cooperation is the basic one. We cannot, I have been at pains to argue, make sense of our lives in terms of this more basic conception. But this is to say something particular about us and not something general.

## Reply to Keren

Keren though sympathetic to the shape of the theory proposed in *Knowledge on Trust* criticizes it in three ways. First, the argument from cooperation has a falsehood as a major premise. Second, the account I give of the psychology of trust cannot explain how belief can be based on trust. Third, I fail to explain how trust provides an *epistemic* reason for belief.

The basic testimonial situation, I argue in *Knowledge on Trust*, presents a problem of cooperation, which can be illustrated by reference to the Trust Game. The problem of cooperation for this game is: why should the investor cooperate and make an initial transfer? There are many answers that can be given to this question and Keren offers several, where that the investor believes the trustee to be kindly is one. What raises the question – why there is a ‘problem’ to be addressed – is that without the further information that explains cooperation, it seems that the game will default to the uncooperative outcome. This is because it would seem to be in the trustee’s interest to keep whatever monies he gains, and so in the investor’s interest not to trust him with any in the first place. The problem of cooperation for the Testimony Game is: why should an audience cooperate and believe what a speaker says? Again there are lots of answers that can be given, where again that the audience believes the speaker to be kindly is one. But the point is that there is a problem at this juncture that needs to be addressed, and this I argue by way of what Keren calls the *parity claim*: “the Testimony Game has the same payoff structure as the Trust Game”. (p.35) This premise Keren believes to be false.

The parity claim is false, Keren contends, because information, and indeed knowledge, “is a good whose consumption is non-rivalrous” (p.36); that is, unlike money, it can be shared without loss. However, it is not the similarity of goods that underlies the similarity in payoff structure or the parity claim. And though it is true that the conflict of interests in the Trust Game derives from the fact that both parties want the same good whose consumption is rivalrous, the conflict of interest in the Testimony Game rather derives from a *difference in interests*. An audience’s basic interest is acquiring a bit of information, and this because of some further need. A speaker’s basic interest is being believed because this is a way of exerting influence, which can similarly further some

particular need. A commitment to sincerity constrains this basic interest, so it is better for a speaker not to be so constrained, and merely to tell the truth when this suits. But this is to be non-cooperative. So while any given conversation will be shaped in determinate ways that can make it cooperative, the basic shape of our conversational engagement is such that without further information that explains cooperation, the game will seem to have the non-cooperative outcome as its default.

Keren's second problem concerns the central claim that testimonial uptake can be based upon affective trust; that trust in this sense can be a reason for belief. Whether trust provides an *epistemic* reason is Keren's third problem, which I will come to shortly, but for now the issue is whether affective trust could be a reason in a "causal-motivational" sense. (p.40) Were it the case that affective trust implied the *belief* that the trusted is trustworthy, there would be no issue, but as I characterize it affective trust merely implies the *presumption* that this is so and, Keren asks, "why should we expect this presumption to result in their believing that the speaker's testimony is true, and not merely in their *presuming* that the testimony is true?" (p.41) A presumption could underwrite belief if it were a presumptive right to believe; that is, if the presumption of trustworthiness were the presumptive right to believe this. However, Keren rightly observes that this is not what is intended, so then he challenges that if the presumption is little more than an assumption made for the sake of argument – as when a judge presumes the defendant innocent – the most it could suffice for would be a state of acceptance. But what speakers expect when they expect to be trusted is to be believed, and trust must suffice for belief if it is to credit this fact and find its place in an epistemology of testimony.

On the other side it is, I think, an important feature of (affective) trust that it does not require a belief in trustworthiness. This is needed if we are able, as I think we are, to choose to trust and to extend trust in the face of doubt. The challenge is then: how can anything short of belief ground testimonial uptake? The idea that trust entails a presumption of trustworthiness was meant to address this challenge, Keren's contention is that it does not. In reply the best I can do is say why I think it does. This presumption is not comparable to that made by the judge, nor is it comparable to an assumption made for the sake of argument because such assumptions are singular and isolated, whereas in the case

of trust the presumption of trustworthiness follows from a set of further acceptances, where these constitute the background set of attitudes necessary for adopting the attitude of trust in the first place. Affective trust is a ‘thick’ ethical concept so that using it is part of living in a particular social world, which is a world where the truth of the set of things accepted in trusting someone is a commonplace. So the presumption is not isolated but a consequence of other things accepted, and these other things, even though they do not need to be believed, will be recognized as things whose truth is commonplace. As such the presumption made in trust might be characterized, if this is any clearer, as a presumptive belief: it is a taking that things are a certain way, but it is a taking that is based on an optimistic view of a world (in that it is both optimistic and expresses a view of the world) rather than a taking based on evidence. So it would be wrong to describe it as a belief, but it has the substance of belief rather than acceptance, which need not be similarly coherent, and it is this that enables it to be the “causal-motivational” basis for belief. This might then be put the other way around: what reflection on cases of trust shows is that we need to refer to an attitude that is belief-like but that falls short of belief and is not constrained in the way that belief is constrained. This is the attitude that I have called a *presumption*. Similarly, that we need such a type of attitude can also be argued by reflection on other cases; for instance Holton argues that we can only make sense of certain cases of intending if we have a notion of *partial belief*.<sup>2</sup>

The third problem Keren finds is with my account of how trust can be an epistemic reason for belief. It is part of my account of how trust can be such a reason that it can be potential evidence. That is, the fact that A trusts S for the truth as to whether p, when S tells A p, can be potential evidence for p; it is so when S is trustworthy. Keren then observes, “even if A believes that p because she trusts S, and her trusting S is potential evidence for p, it is not at all clear that it would be correct to say of her belief that it is based on this evidence” (p.44). However, potential evidence for p can justify A believing that p if and only if A believes that p on the basis of this evidence. Since this condition is not satisfied

---

<sup>2</sup> See Holton (2009), pp. 29-34. Though the presumption of trustworthiness is not a partial belief as a key feature of this presumption is that the trusted party’s being untrustworthy is not a “live possibility”.

when testimonial uptake is based on trust, this account cannot explain how it is that trust provides an epistemic reason for belief in these cases.

If the warrant that trust provides comes by way of trust being a piece of evidence, then this is correct: the audience's belief needs to be based on trust *qua* evidence, and it is not so. But the warrant that trust provides does not come this way, rather trust provides a way in which an audience's belief can be based on the extended body of warrant possessed by the speaker (and testimonial chain) and it is this that warrants the audience's belief. However, in order to play this role of basing the audience's belief on this extended body of warrant, trust needs to make testimonial uptake epistemically reasonable. The argument that Keren considers is then addressed to this question: not how does trust warrant but can trust be considered to be an epistemic reason? The argument is this: trust provides a "causal-motivational" reason in that when A confronts S's telling him p, A's trusting S for the truth makes it subjectively probable for A that p; and this reason is an epistemic reason because A's trust could be evidence for p. Thus, it is not that trust provides the reason of evidence but that the reason that trust provides is an epistemic reason because trust could figure in an argument for the truth of that which it is a reason for, because it could be a piece of potential evidence.

### **Reply to Hinchman**

There are, I think, essentially three criticisms that Hinchman directs against the theory of testimony put forward in *Knowledge on Trust*. The account I offer of trust-based reasons is implausible. The solution I offer to the problem of cooperation is incoherent. And the analysis I give of affective trust fails of sufficiency.

First, affective trust, as I conceive it, provides a reason for belief because it is essential to trusting that one presume the trusted will behave as one expects; that is, and in short, because one presumes the trusted to be trustworthy. Take the case of an audience A trusting a speaker S to tell the truth. As Hinchman observes, it is consistent with A trusting S in this respect that S be quite untrustworthy. From this observation Hinchman then infers that "trust cannot itself provide a reason to believe the trusted, independently of the speaker's status as reliable". (p.56) And since trustworthiness implies reliability – merely

saying what one believes is not enough – the inference is: the trusted has a reason to believe *only if* the presumptions made in trust are true. This condition is necessary because “without an appeal to S’s reliability, an account of A’s testimonial reason would admit the possibility that A has bootstrapped his way into possession of a reason through his mere affective trust in S.” (p.55). And this, Hinchman thinks, is implausible.

Now the truth of what is presumed in trust matters. It matters to A’s epistemic standing: whether trust results in knowledge or warranted belief or not, will hinge on whether S is trustworthy. And it matters to the status of A’s reason for belief: were A’s presumptions true, that A’s trusts would be a *good* reason for belief in the sense that it could be used by a third-party in a justification for p, when this is what S tells A; it would be a bit of potential evidence for p. But the truth of what is presumed in trust is *not necessary* for trust being a reason for belief for A. Rather, trust is a reason because in trusting S for the truth, A’s attitude makes it subjectively probable for A that p, when this is what S tells A, and our social world is such that A’s trust *can be* a good reason in the sense noted. In allowing that trust can provide a reason for belief *and* allowing that trust can be chosen, it follows that A can bootstrap his way into the possession of a reason. But this conclusion is not implausible given its presuppositions, namely that this possibility requires as a background the social institutions of trusting and telling.

Second, the solution I offer to the problem of cooperation, Hinchman argues, is incoherent in that it requires both an internalist and an externalist conception of practical reasons. What is presumed in trust is that the trusted will see things in a certain way, and as a result have certain motivations. In trusting S for the truth as to whether p, A presumes that S will be motivated to tell A the truth by A’s need for it. Here reasons are conceived in internalist terms. However, and here Hinchman asserts the necessary condition just disputed: A possesses a trust-based reason *only if* S does in fact have this reason to tell A the truth. “Faulkner’s strategy”, Hinchman continues, “is to turn that necessary condition on trust-based testimonial reasons into a sufficient condition.” (p.57). Genealogy accomplishes this transformation: if we have escaped the State of Nature, social conditions must be in place such that *anyone* in S’s position would have a reason to tell A the truth. Thus trust is sufficient for reasonable belief and the problem of cooperation is resolved.

However, this resolution presupposes an externalist conception of reasons – it presupposes that S has a reason merely because of the position that S is in. Given that both internalist and externalist conceptions of reason are presumed, the result is incoherence.

It is true that it is the availability of trust-based reasons that resolves the problem of cooperation, and that what makes these reasons available is that certain social conditions are in place, but the genealogy does not thereby establish their availability through showing that *everyone* has the reason to be truthful that it is presumed the trusted has in trust. As Hinchman observes, it is not true that everyone would care about the trusting party's needs, and there are those, who I follow Hinchman in calling 'psychopaths', who could not be brought to care. However, the argument is not that even psychopaths have a reason – an externalist reason – because what makes trust-based reasons available is not that the putative condition on their possession is universally satisfied; rather, it is simply that the social conditions are such that trust generally, though not always, is an option. Its being an option is what the genealogy establishes, and it is not always an option because one might, for instance, believe the other party to be a psychopath.

If there are norms of trust, then it is possible to go wrong in one's thinking about the trust situation. And in terms of these norms the psychopath does go wrong. However, to suppose that any failure to think about the trust situation in the prescribed way is a failure to be moved by *reason* requires supposing that rational deliberation in conformity with the norms of trust is just a matter of reasoning as the fully rational person would. Here I agree with Williams: accusations of irrationality are mere "bluff".<sup>3</sup> It seems reasonable, for instance, to think about the trust situation in terms of the logic of tit-for-tat even if to operate in accordance with this strategy is to demonstrate little trust (in the affective sense). We think about things in terms of the norm, this is how we have resolved the problem of cooperation confronted in the State of Nature, but the psychopath has slipped through the net. Thus a short response to this criticism is simply that the only conception of practical reason in play is internalist (and Hinchman only thinks otherwise because he takes the disputed necessary condition to hold).

---

<sup>3</sup> Williams (1980), p.111.

Third, the analysis I give of affective trust fails of sufficiency because it aims to capture the distinctively second-personal character of trust as it can be found in the testimonial situation and yet the conditions proposed can be satisfied by trust as it is found in *institutional settings*, which are not second personal in any way. Hinchman gives the example of A (an account holder) trusting S (a bank teller) to transfer some monies. In this case, the dependence condition is satisfied: A depends on S transferring these monies. And the expectation condition is satisfied: A expects S to transfer the monies precisely because this is what he, A, depends on S doing. Moreover, this expectation is normative: in the given situation, A thinks that this is what S *ought* to do and this is what A *expects of* S. However, the normativity here is not second-personal but is mediated by the institutional setting. It rests on no more than the thought that S should do her job. Since the conditions on affective trust are met but the trust is not second-personal, these conditions do not suffice to give an account of this dimension of trust.

This is an interesting case but, on balance, I think the best response is that it is not a case of affective trust, and so not a counterexample to the use of this notion of trust in giving a theory of testimony. It is *not* that A (the account holder) expects S (the bank teller) to respond to *his*, A's, dependence but that A expects S to respond to the demands of account holders generally. This is what is involved in S doing her job: *not* responding to the needs of particular individuals but responding to the demands of individuals *qua* account holders. So what the case combines is predictive trust with normativity. It combines a belief that someone will act in certain ways with the belief that acting in these ways is prescribed. Here it is worth stressing that our reason for testimonial uptake can often be such a straightforward assessment of the likelihood truth and that such an assessment can be, as in this case, straightforwardly based on assessments of motivation, which in turn can revolve around judgements about social and institutional norms. Testimonial trust does not always have the second-personal character that I hope to capture by reference to affective trust; it can be simply predictive. The philosophical point is then that the success of this kind of empirical judgement presupposes large-scale consistencies in motivations, which are secured by the norms of trust (even if these norms do not shape motivations universally and there remain psychopaths). So, in the genealogical story, assurance comes before evidence.

### Reply to Hawley

Like Keren, Hawley thinks that the argument from cooperation is based upon a falsehood. The falsehood is that the Testimony Game is analogous to the Trust Game. Hawley argues that it is not, and what follows from this is that I have failed to establish that the uptake of testimony requires a reason if it is to be reasonable. That is, what follows is that the argument from cooperation fails to conclude. In reply I would argue that the analogy between the Testimony Game and the Trust Game amounts to a specific claim: these games have the same pay-off structure. This I argue on the basis of assumptions about the basic interest each player has in the game, where these assumptions are defeasible and their defeat would then constitute a specific reason for expecting a cooperative outcome. In arguing that the games are disanalogous Hawley does not question these assumptions or that the games thereby have the same pay-off structure. Rather, she argues that the games are disanalogous on the basis of a number of differences between them. Of course, that the games are analogous in the sense specified is consistent with their being different in other ways. So the question is: *are the differences Hawley highlights such that the similarity between the games with respect to pay-off structure cannot be used as a premise in the argument from cooperation?* In what follows I argue that the differences Hawley highlights are not as significant as she suggests, and that these differences do not undermine the argument from cooperation.

There are, Hawley argues, four differences between the Trust Game and the Testimony Game: there is a temporal difference in the order of play; there is a difference in the strategies that can be adopted; there is a difference in the currency of rewards; and the roles played in each game differ in their stability. However, Hawley only uses the first two of these differences in arguing that the games are disanalogous in a way that undermines the argument from cooperation, so I focus only on these two.

First, there is a temporal difference in the order of play. In the Trust Game the player who must decide whether to trust (the investor) makes the first move in making a transfer of monies or not, whereas in the Testimony Game the player who must decide whether to trust (the audience) makes the second move in that this decision is a response to a speaker's assertion of something. This undermines the argument from cooperation,

Hawley claims, because the very fact that a speaker has made an assertion can be taken as a reason for thinking that the speaker is cooperating.

That there is a difference here, I think, must be acknowledged, but it is less than it seems and does not undermine the argument from cooperation. In the Testimony Game, a speaker's assertion purports to address an audience's need for information, say as to whether *p*, and can, as such, be regarded as a response to this recognized need. That is, the audience's need is regarded by the speaker as posing the question, *p*? Were this question vocalized by the audience, the difference Hawley highlights would vanish. But then does this vocalization really make such a difference? That a speaker's assertion comes first undermines the argument from cooperation only if this assertion in itself constitutes a reason or thinking that the speaker is being cooperative. Later in *Knowledge on Trust* I offer one account of how this might be so in outlining the assurance view, and endorse something like this view. But the problem for this view, and indeed for any suggestion that a speaker's assertion itself offers a reason, is that multiple explanations can be given of utterance, including, in particular, explanations that do not start from a speaker having any informative intention. This account denies the assumption about interest made in setting up the Testimony Game, but if this assumption is accepted, and Hawley does not question it, then the temporal order of play makes no difference.

Second, there is a difference in the strategies that can be adopted by the players in the two games. In the Trust Game the player who must decide to trust (the investor) can adopt the strategy of "never take a risk" – or "never transfer cash" – and so accumulate capital, whereas in the Testimony Game if the player who must decide to trust (the audience) were to adopt the strategy of "never take a risk" – which in this case amounts to "never give credence" – the result would be unremitting ignorance. This undermines the argument from cooperation, Hawley suggests, because it implies that uptake must be default entitled.

Again I think that there is a difference here but it is less than it seems and it does not undermine the argument from cooperation. It is true that the non-cooperative strategy is not a bad strategy for the investor in the Trust Game in that by following it he might accumulate cash. But it is still a worse strategy than cooperation when this is successful; the

investor might get rich through taking no risks but could get much richer by taking risks. And it is true that the non-cooperative strategy is not a good strategy for the audience in the Testimony Game in that by following it he is simply left in a position of ignorance. But it is still a better strategy than cooperation when this is unsuccessful; the audience might remain ignorant through taking no risks but could end up in error by taking risks. It is better to give no credence than give credence wrongly just as it is better to lose no money than lose all one has. So the difference is not so great. And rather than undermine the argument from cooperation all that is implied by the fact that we are not ignorant is that the problem of cooperation is empirically resolvable. Our risk taking is often, even ordinarily, well judged in that we tend, by and large, to give credence when credence is due. What I then argue in *Knowledge on Trust* is that this (reductive) solution to the problem of cooperation is limited in various ways; in particular, it is limited in that it accepts that there is a problem confronted. A more radical solution involves denying the background assumption, which leads to the problem, that competing explanations of utterances are equal. But there is no short step from noting the sceptical nature of the problem of cooperation to such a denial. In particular, our possession of an entitlement to believe testimony would require some basic connection between testimony and truth but if the assumptions about interest that generate the problem are correct any such connection is contingent.

### **Reply to Graham**

I would like to make three points in reply to Graham. First, I would like to try and clarify the psychology of trust and the role that trust plays in the epistemology I propose because Graham does not get these quite right and his not doing so feeds into his first, and principle, objection. And then, second and third, I will try and respond to Graham's two objections: that there can be no epistemological role for trust once social norms of trust are recognized; and that my theory of testimony is refuted by the fact that children can acquire testimonial knowledge.

As Graham reports my account, the psychology of trust involves the trusting party having a number of beliefs and making a number of presumptions. For the case where A trusts S to  $\phi$ , trust involves A believing that S can recognise his, A's, dependence and

believing that S can recognise that his, A's, attitude towards this dependence is one of trust. And trust involves A presuming (i) "that if S recognizes A's expectation that S should prove informative, then other things being equal, S will prove informative for this reason"; and so presuming (ii) "that the trusted will prove trustworthy". (p.99) This description gets the belief side right, and trust does involve presumption (ii), but this presumption does not rest on (i), which is no part of trust. What is presumed is that S will be moved by the *fact of* A's dependence, which A believes S recognises, not by any normative expectation that A might have. To see this suppose that S is trust-responsive in Pettit's sense and is particularly sensitive to A's opinion and desirous to avoid A's resentment. If this were the sole reason that S  $\phi$ -ed, then presumption (i) would be true and yet presumption (ii) false because S would not have the motivations characteristic of the trustworthy person. So (i) could not be the basis for concluding (ii) and, indeed, it is not so.

This is then relevant to the main criticism Graham articulates because he finds two roles for trust in my theory – a rationalizing role and a motivating role – and his criticism is that once social norms are recognised both roles are epistemologically otiose. The motivating role is meant to be that just described. The trusting party A is meant to presume (i) and this presumption is meant to come out true such that S is motivated to  $\phi$  by A's trusting S to do this. This can be the case: S can be trust-responsive, and so A's trust can give S an instrumental reason to behave as A expects. However, this reason is not the reason that would move the trustworthy person, who would be moved simply by A's need, rather than by A's attitudes. So trust does not play the motivating role Graham describes. So it can be no criticism that this role is epistemologically superfluous. However, trust is meant to play a rationalizing role, so Graham's main criticism still needs to be addressed. I turn to this now.

In *Knowledge on Trust* I argue that the problem of cooperation establishes the requirement that the uptake of any piece of testimony needs to be rationally supported. And I argue that the attitude of (affective) trust can play this role: it can be reasonable to base belief on trust. Graham's criticism is that the existence of a social norm of truth-telling establishes an entitlement to believe testimony. So if there is such a norm, there can be no requirement that testimonial uptake be supported in every case. So there is no essential

rationalizing role for trust to play. Of course, our entitlement to testimony might be defeated in any particular case, so the uptake of a piece of testimony might require supporting reasons and trust could then play a rationalizing role. But this role is non-essential and what matters is the fact that speakers are by and large reliable, where their being so is established by the social norm of truth-telling; “[s]peaker reliability, grounded in pro-social preferences, is enough.” (p.115)

Suppose it were established that there is a universal norm of truth-telling. (For instance by appeal, following Longworth’s suggestion, to a stronger conception of practical reason than that which underpins the problem of cooperation.) If this were established, then there would be a basis for arguing that we have an entitlement to believe testimony. However, the claim that there is a social norm of (informative) truth-telling is not this universal claim. It is the claim that there is a *social* norm, a norm that operates in this vicinity or for people like us. By contrast, the requirement established by the problem of cooperation is meant to be universal. This is because it is an essential feature of communication that speaker and audience have different basic interests in the communicative exchange. So any instance of testimonial uptake must be rationally supported. What the institution of social norms of trust then establishes is one way in which this requirement can be met: in this locality, uptake can be based on trust. It might then be claimed that what is thereby established is that *we* have an entitlement to believe testimony, where the referent of ‘we’ ranges over the domain of the social norm. However, this would be an entitlement in name only because it would be grounded on a particular empirical claim, and so, in fact, would be a reductive justification with a major premise of the form ‘speakers of type X tend to be reliable’. This gives a straight solution to the problem of cooperation; but this is not the solution proposed in *Knowledge on Trust*, which is rather that where there are social norms of trust interlocutors can operate with a series of presumptions that make trust non-problematic. This dissolves the problem not, as Graham suggests, through adding an additional set of pro-social preferences, but through denying the conception of rationality presupposed in setting up the problem of cooperation by widening the class of things that can act as a reason for action. Thus reasons are not limited to complexes of belief and desire but can include the trustworthy speaker’s perception of

the audience's informational need. However, this dissolution travels only as widely as the norms of trust are internalized, and these are merely local distances.

The second objection put by Graham is that facts about children refute the Trust Theory of testimony, and my argument otherwise fails. Specifically the following facts form an inconsistent set.

1. An audience A is warranted in the uptake of testimony to p if and only if it is reasonable in the light of A's other attitudes to believe that p.
2. The attitudes that can render the uptake of trust reasonable are either A's beliefs, or A's trust.
3. Children do not possess enough background of belief to render uptake reasonable.
4. Children do not have the sophistication needed for adopting an attitude of (affective) trust.
5. "Nonetheless, children learn (come to know) a great deal by believing what their parents, relatives, teachers, caregivers and even strangers tell them."  
(p.112)

In *Knowledge on Trust* I considered this argument and proposed that neither 3 nor 5 were clear cut. At some point, children gain sufficient sophistication and belief for uptake to be reasonable in the light of it, and at some point children come to know things on the basis of testimony. But at no point is it clear, as the argument requires it to be, that children lack sufficient sophistication and belief for uptake to be reasonable but nevertheless come to know things on the basis of testimony. Graham rejects this response to the argument. First, what I cited as evidence of the sophistication of children, namely their sensitivity to a speaker's communicative intentions, is not sufficient for the possession of a reason. Second, denying that children acquire testimonial knowledge amounts to just "denying the data".  
(p.113)

In response, what I cited as evidence that children are more sophisticated than proposers of the argument would assert in asserting 3 was in fact an article by Paul Harris. I now quote from this.

I resist the claim that credulity is strongest in early childhood. Young children are endowed with two protective devices. First, they are alert to a speaker's intention – they do not systematically confuse fictional and factual claims. Second, for comprehension to proceed smoothly, any new piece of testimony needs to be consistent with, and integrated into, what is already known about the topic in question. Hence we may assume that young children will find it difficult to accept and integrate what they regard as anomalous statements. Indeed, to the extent that they know less than adults, they may ultimately be less credulous. Their impoverished knowledge base will make the integrative process slower or more taxing. Stated simply, children's ignorance may often safeguard them from misplaced trust.<sup>4</sup>

If Harris is correct, then testimonial uptake can be reasonable even for young children: testimony which is believed is believed because it fits with the child's background of belief. Now I do not want to assert this consequent, but I think that this possibility is enough to establish that 3 is not clear cut, as claimed. What, then, of the 'data'? It cannot be that children learn things from testimony. This is certainly true but it is so because it can be interpreted as no more than the claim that children get to form true beliefs on the basis of testimony. What is needed is the claim that children acquire *knowledge* in some stronger sense. And there is just no clear cut data here as to when this happens in part because of disagreements as what knowledge in this stronger sense amounts to. However, if one takes knowledge to be a certain 'standing in the space of reasons', as I do but Graham doesn't, then it would be natural to regard the shift from 3 being true to being false that occurs with a child gaining belief and sophistication as tracking the shift from 5 being false to being true. But if this is the case, there is no point at which 3 and 5 are both true, as the argument asserts.

---

<sup>4</sup> Harris (2002), p.331.

**Paul Faulkner**

*University of Sheffield*

paul.faulkner@sheffield.ac.uk

## **References**

- Harris, P.L. (2002) 'Checking our sources: the origins of trust in testimony'. *Studies in the History and Philosophy of Science*, 33:315-333.
- Holton, Richard (2009) *Willing, Wanting, Waiting*. Oxford: Clarendon Press.
- Williams, B. (1980) 'Internal and External Reasons'. In *Moral Luck*, edited by B. Williams. Cambridge: Cambridge University Press.