

UNIVERSITY OF CALIFORNIA

Los Angeles

An Ethics of Uncertainty:  
Moral Disagreement and Moral Humility

A dissertation submitted in partial satisfaction of the  
requirements for the degree Doctor of Philosophy  
in Philosophy

by

Christopher Thi Nguyen

2011

# Chapter 1

## Moral Disagreement and Moral Humility

We are fallible beings. All of our calculations, judgments, and decisions rest on methods and abilities that are not, and cannot be, perfectly secured against mistake.

Though we must trust our mental abilities in order to think and act, that trust is vulnerable to revision. The inescapable fallibility of human cognition is one of the major conclusions of modern epistemology. Yet we often ignore the fallibility of human cognition in moral reasoning. Many of us - academics and laypeople both - take moral judgments to be not only expressions of our inner feelings, but judgments about some objective fact. Moral reasoning is as fallible as any other cognitive domain. Yet we often behave as if moral judgment were infallible.

This is, I claim, epistemically naïve. I argue for a more reasonable stance towards our own moral intuitions and judgments. I argue neither for total skepticism towards our moral judgments, nor unqualified acceptance, but for tentative acceptance, laced with a goodly amount of suspicion. I argue for *moral humility*.

For this dissertation, I will focus on the interaction of moral disagreement and moral self-confidence. It has often been thought that autonomous agents should not let moral disagreement change our beliefs. There seems to be something special about the moral domain, something that demands that we make up our own minds about the matter. We ought not bow down to moral authority; we ought not acquire moral beliefs from testimony; and we ought not change our beliefs from the mere existence of moral

disagreement. So, goes this line of thinking, we ought to reject all social sources of moral knowledge. We ought to strive for moral judgments that proceed from our own understanding, that gel with our genuine and sincere moral intuitions. Though conversations with others might yield new arguments or reasons for us to consider, in the end only we can make up our mind for ourselves. We should only believe according to those reasons and arguments that seem to us to be correct. We should be morally self-sufficient.

I argue that this claim of the essential self-sufficiency of moral reasoning is flawed, and rests on naïve epistemology. Once we bring more sophisticated epistemology to bear - once we try to take seriously what it's like to be the sorts of beings that must always reason with unsecured foundations in every domain - we will see that we must allow certain types of disagreement to have weight in any cognitive domains, including the ethical. We are cognitively fallible in virtually every domain. Though it is important that every agent conduct independent moral reasoning, that requirement is not the only relevant consideration. We cannot ignore our fallibility; it is our fallibility that forces us to attend to moral disagreement.

The claim of absolute moral self-sufficiency is an exaggeration of a more plausible consideration, that there is something wrong with giving ourselves over entirely to the command of another. Though I grant that there is something wrong with complete moral obedience, I claim that the use of disagreement I will describe does not count as obedience. It is, instead, a distinct process, part of an epistemic procedure of self-checking and corroborating our own mental abilities. Employing this procedure does not

count as obedience; it is, in fact, virtually the antithesis. The drive to self-checking comes from the very same values that lead us to abhor moral obedience: a drive to understanding, to self-perfection, to greater responsibility for our moral beliefs; a drive to get our morality right. Moral obedience is a form of disengagement from the moral process. Using social sources of information to corroborate and dis corroborate our cognitive abilities is actually a form of increased engagement with moral reasoning, and a part of mature rational approach to moral judgment.

My central claim is that moral disagreement, in and of itself, matters. I am not merely making the weaker claim that moral testimony and moral disagreement can give us reason to *reconsider* our views. Many hold that weaker view, and yet go on to say that once I have re-considered the claims and settled my own mind about the matter, then I may safely disregard the disagreement. That view strikes me as still essentially within the camp of moral self-sufficiency; it treats my own settled judgment as definitive. My stronger claim is that the presence of peer disagreement in and of itself gives us reason to *doubt* ourselves - to doubt our beliefs, and to doubt the faculties and methodologies that generated those beliefs. Even in those situations where I've considered my opponent's arguments and find them unconvincing, and re-considered my arguments and found them sound, the mere presence of a rationally and morally respectable thinker who disagrees with me gives me a reason to doubt myself, even if I cannot fathom their thinking.

I am not claiming that all moral disagreement matters; surely disagreements with some sorts of intellectually or morally disreputable sources can be safely discarded. I am arguing that some types of moral disagreements - especially those with particularly

trustworthy and reliable individuals matter; I am arguing that moral disagreement is the *type of thing* that can matter. Just as in non-moral domains, there will be instances of really compelling disagreements - disagreements with peers and experts – and instances of entirely unconvincing disagreements. Moral disagreement is of a piece with disagreement in any other cognitive domain, I argue, and important in precisely the way that scientific disagreement is important. The existence of moral disagreement, especially long-standing intractable disagreement, with our intellectual and moral peers, gives us reason to doubt those moral beliefs over which we disagree. Disagreement ought to lead to self-doubt, and self-doubt ought to lead to restraint in action. This is the form of moral humility I will begin to argue for.

This argument for moral humility is part of a larger project, which is to characterize the epistemically unstable situation we find ourselves in. We have different commitments - commitments to epistemic standards for objective inquiry, moral commitments, commitments to having independent moral judgment - and these different commitments do not happily co-exist. We cannot help but trust ourselves, and we cannot do without trusting other people, but these trusts lead to conflict. Where we end up is in an unsettled, uncertain position - but one from which we are still called to act. How are we supposed to conduct ourselves from a position of moral uncertainty? This, I think, is the key question, and one that has been radically unexplored. I cannot hope to settle that larger question here; my only hope is to begin the project, and begin to show that our moral lives fall under a cloud of doubt. To do this, I will show that we cannot reasonably dismiss the testimony of certain other people, and as a result, the existence of unresolved

disagreements with them ought lead us to reduce our self-confidence. Moral self-sufficiency is wrong; we must trust others and they will render us morally humble.

## **Part I. Moral self-sufficiency**

Our lives are filled with moral disagreement. We disagree not only with psychopaths and madmen, but often with reasonable, moral people. I disagree with intellectuals and philosophers from other cultures about many basic moral values; I disagree with political opponents in my own culture about crucial issues of justice, punishment, freedom, and sacrifice. I disagree with my political allies about the relative importance of issues about the environment, social welfare, and defense. I disagree with my parents about the importance of tradition in my life. I disagree with my close and trusted friends about the importance of animal welfare, the importance of honesty. Disagreement is frequent, even with people we trust, admire, and respect.

If this is so, and if I am right that moral disagreement matters, then we all ought to be plagued with moral self-doubt. Instead, for most people, academic and non-academic alike, moral life is marked by self-confidence. Often, the people we praise the most - our political and moral heroes - are those who are the least morally humble. We praise them as men and women of moral conviction, of commitment, integrity, and unswerving devotion to their moral ideals. In the methodology of much academic ethics, there is also a prevailing attitude of moral self-confidence. Academic ethics often proceeds with a mood of self-confidence. We often presume that we are justified in trusting our moral

intuitions, and we accept theories when they match our intuitions. In order for this self-confidence to exist against the background of plentiful disagreement, one must deny the importance of moral disagreement. Thus, most people seem committed to some form of moral self-sufficiency.

Why do we so often deny the effects of social corroboration and discorroboration in the moral sphere? Corroboration and discorroboration are crucial in most other cognitive spheres, especially with empirical judgments. Agreement, we think, means something in checking scientific laboratory results, checking mathematical calculation, and checking old memories. In most cognition, corroboration gives us some reason to trust our mental abilities more, and discorroboration gives us some reason to reduce our mental self-trust. What most of our behavior seems to encode, then, is the belief that moral judgment is an exception to these methodologies, that it is not open to the usual social methods of self-assessment.

It can seem quite plausible to grant moral judgment such an exception. Agents seem to have a very special, peculiarly personal relationship to their moral beliefs. Consider, for example, the widespread intuition there is something wrong with moral *deference* - the wholesale acquisition of new moral beliefs solely through testimony. There is no analogous block to deference in empirical reasoning. There, deference to experts seems not only unproblematic, but requisite. The sheer intellectual complexity of contemporary life demands that we trust the empirical judgment of others without, in many cases, understanding the basis for that judgment for ourselves. We may master one or two domains, but no single person can adequately understand modern medicine,

nutrition, automotive repair, computer software architecture, and meteorology. We must trust and act on the judgment of others. That trust is not necessarily blind; we may only trust that we have some reason to think of as reliable), but we must trust. We may be able to give reasons why we trust this person, but sometimes we must adopt some particular belief without understanding the direct evidential grounds for that particular belief ourselves.

But things seem very different with moral beliefs. Though we don't hold every person responsible for understanding their medical beliefs, we do seem to hold that every person is responsible for arriving at their own moral beliefs independently, through their own reasoning and understanding. The requirement of independence for moral judgment can seem quite plausible and desirable. We prize moral thoughtfulness; we respect those that take responsibility for their moral judgments, who think through their beliefs. We find unsavory those who acquire most of their moral beliefs unthinkingly, by trusting others. Excessive moral trust is seen at best as naïve and childlike, and, at worse, as evil; think of that archetypical villain of moral mediocrity: the unthinking soldier, only following orders. The practice of moral judgment seems to be the responsibility of every full person.

There are at least two distinct versions of the independence requirement that might threaten my claims about the importance of disagreement. The primitive thesis is that any and all uses of moral testimony are illegitimate - that the process of moral judgment ought to be entirely private. Call this the thesis of *moral isolationism*. The more sophisticated thesis is that the use of moral testimony ought to be carefully



circumscribed, and that, in the end, one ought to possess and accept the justification for one's own moral beliefs. Moral testimony as to a certain belief can give one a reason to consider that belief; moral disagreement can give one a reason to go back and re-consider one's own moral beliefs. But, once one has re-considered, the epistemic use of other people has exhausted itself. Other people may point out reasons to us for our own inspection, but one's moral judgment ought to proceed from reasons one accepts and holds, by reasoning one accepts and understands. Let me call this more sophisticated thesis the thesis of *moral self-containment*, since the thesis demands any moral belief I hold must be justified on reasoning I myself possess, understand, and endorse, even if I had social help acquiring that justification. Under this thesis, I cannot justify a belief by pointing outwards to another's beliefs. Collectively, I will call these theses the theses of *moral self-sufficiency*. Both forms of moral self-sufficiency severely limit the role moral discourse. Others may point out reasons to us, introduce us to arguments, and help us to understand them. But when we finally make up our own minds, it must be based on reasons that we ourselves find compelling, by arguments that we are convinced by. The elucidation of arguments and reasons exhausts the role of moral discourse; the mere fact that there are still opponents who disagree does not have any weight on its own.

Either version of moral self-sufficiency would block my claim of moral humility. If disagreement is to have its fullest significance, then we must treat the bare fact that somebody disagrees with us as important, even if we've settled and re-settled our judgment about the matter. Both versions of moral self-sufficiency do not permit weighting the bare fact of disagreement. Under moral isolationism, we do not even allow

the testimony of others to open any questions about our moral judgment. Under moral self-containment, we do allow disagreement to raise a question about our moral judgment, but we take our own process of re-inspection as settling the matter. Under either thesis, the mere fact of disagreement cannot impugn our cognitive self-trust; if disagreement raises a question about the reliability of our cognitive processes, we treat our own cognitive processes as either unimpeachable in the first place, or able to satisfactorily settle the question.

My goal in this dissertation is to argue against moral self-sufficiency. While I readily admit that outright moral deference is problematic, I will claim that this problem is confined to particularly egregious cases of obedience, and does not apply to most uses of disagreement. Most uses of disagreement are not forms of obedience, but rather part of a rational and autonomous process of cognitive self-perfection. We are not cognitively self-sufficient in the moral realm any more than in any realm; we can and should use others to corroborate and dis corroborate our moral judgments, and when we find particularly pernicious forms of moral disagreement, we should become suspicious towards our own moral judgments as well as towards others' judgments. The fact that we have double-checked our moral reasoning cannot completely settle the matter, for disagreement raises a question about the reliability of tools that are used in all forms of moral reasoning.

I will argue that the view of moral self-sufficiency cannot survive a sophisticated appraisal of the epistemic status moral judgment. As long as we take ourselves to be cognitively fallible beings, and as long as part of our moral enterprise is to *get it right*,

then we must be open to the evidence provided by disagreement. I will argue that any commitment to the objectivity of moral judgment will involve a commitment to certain basic epistemic principles, which entail limitations on our self-trust. I hope that this epistemic analysis will legitimize social methods of corroboration and discreditation for moral judgment. Insofar as we take our moral judgments to be cognitive, then unresolved moral disagreement ought to lead to self-doubt and moral humility. Moral self-sufficiency is epistemically unreasonable.

My argument rests on the epistemic analysis of moral judgment. I will spend the majority of this dissertation arguing: first, that a reasonable epistemic grounding for moral judgment admits of fallibility; second, that moral disagreement can provide positive evidence of cognitive unreliability; and third, that nothing about the moral domain can exclude these effects from disagreement.

## **Part II. A new epistemic foundation for moral judgment**

The commitment to moral self-sufficiency rests on a naïve epistemic understanding of moral judgment. Moral epistemology has been historically underdeveloped; for much of the history of metaethical thinking, we have not had any epistemic theory capable of adequately accounting for the seeming peculiarities of moral judgment. But recent developments in epistemology promise to improve matters. I will show that moral judgment is not as epistemically peculiar as was once thought, and that our moral judgments are in principle just as reasonable to trust, and just as fallible, as our

abilities in any other cognitive domain. This new epistemic grounding will show that moral self-trust is reasonable, but only if it is defeasible; that defeasibility will be the lynchpin in overthrowing moral self-sufficiency.

Our moral judgments are complex, substantive, and often embarrassingly underjustified. Ordinary moral judgment often involves applying difficult moral concepts to complex, fluctuating states of affairs. In the course of a day's work, we must trust substantive raw moral intuitions about particular situations; we must also trust our ability to quickly apply complex moral principles to particular situations. In the past, no epistemic theory could account for the degree of self-trust needed for any reasonable moral life. As long as our available epistemic theories demand that all our judgments and cognitive abilities be fully justified, everyday moral judgments seemed indefensible. Only two reactions seemed possible: accept that we needed but couldn't provide sufficient justification for our moral judgment, and, as a consequence, abandon trust in our moral judgments entirely; or refuse to abandon our moral judgments, and abandon instead any requirement for normal epistemic justification in the moral sphere. New work in epistemology, however, suggests a new and better option, which will provide a more palatable middle path.

## Unsecured moral judgments

The subject of my discussion is what I will call *unsecured moral judgments*. An unsecured judgment is one whose conclusion is not secured by proof from self-evident grounds. The category includes what are often called raw moral intuitions - substantive judgments about particular cases, based on phenomenally direct apprehension of seeming moral properties. The judgment, for instance, that one ought not push a fat man from a bridge onto the train tracks to divert a trolley, believed because it just seems obviously so, counts as a raw moral intuition. The category of unsecured judgments also includes conclusions deduced from unsecured or intuitively held principles. For instance, if we rigorously deduce from a utilitarian principle that we ought to kill one person to save one million, but we hold the utilitarian principle simply because it seems right then that particular judgment is also unsecured.

The category of unsecured judgments also includes conclusions drawn from self-evident principles by non-deductive means; call these *rough applications*. For example, though it might be self-evidently true that three inches is longer than two inches, an everyday judgment that this flower over here, which looks about three inches long, is longer than that flower over there, which looks about two inches long, still counts as unsecured. Even though the central principle may be self-evident, the process of applying that principle to the world depends on the application of a fallible cognitive ability to connect the principle with a particular state of affairs. Since the cognitive ability is unsecured, all judgments that rely on the ability are also unsecured.

Rough applications are very common in daily, non-philosophical life. Often we must make and trust judgments, in limited time-spans and with limited intellectual resources, about time, distance, difficulty, people's facial expressions, human psychological responses, and other complex causal interactions. Much of practical life proceeds from very rough judgments - judgments based on some synthesis of analysis, principles, heuristics, past experience, and native ability. These are judgments often performed at speed, with insufficient information, where a multitude of principles, considerations and values entangle, with no clear singular decision procedure. Examples include judgments that another person is happy, that they're hiding something, that they're unreliable behind the wheel, and that it feels like it's going to rain soon.

Most of the moral judgments we actually make are unsecured. In our present theoretical state, moral judgments often involve either direct raw intuitions or unsecured applications of principles, or often both. Furthermore, many of principles invoked in unsecured applications are themselves also unsecured; they are either held on an intuition, or justified with ineliminable references to unsecured intuitions.

Unsecured moral judgments are very important to consider for a number of reasons. First, I do not think ethical theory possesses at present the resources to secure moral judgment; I am beginning to suspect that ethical theory will never produce anything like entirely secure foundations for moral reasoning. Thus, our moral judgments are surely unsecured at present, and will likely remain so. Second, even if we do produce secure foundations for some ethical principles, given the finite cognitive abilities and time of humans, many of the applications of those principles will be rough, and thus

unsecured. Even though we might have a secured math and secured geometry, action in the physical world requires the use of unsecured abilities to apply these principles to rapidly fluctuating states of affairs. Similarly, even if we could provide a secure proof for, say, the principle that we should promote the ends of others as vigorously as our own, the application of that principle would require the synthesis of such a vast array of facts and considerations that the application of that principle to the world must be unsecured for finite beings like ourselves.

Third, it may be that moral judgment is *essentially unsecurable*. Warren Quinn argues that if forced to choose between accepting a philosophical argument for skepticism about the external world and believing in a simple external object like a chair, he would take the chair.<sup>1</sup> Quinn's argument is not simple-minded; rather, it is a reminder that reasoning itself is a defeasible process, substantively dependent on fallible cognitive faculties like short-term memory, deduction, and the like. Both processes - the visual process by which I see the chair and the reasoning process that casts doubt on the chair's existence - are defeasible. I take Quinn to be suggesting that, once we see that both sides of the conflict depend on defeasible trust, it's quite reasonable to stand with one's visual abilities against one's philosophical abilities. After all, for most of us, we've made far more philosophical mistakes than mistakes about nearby middle-sized objects. But if, in a given domain, our confidence in our unsecured entry-level facts, be they perceptions or intuitions, exceeds our confidence in our ability to secure those facts through some reasoning process, then the domain will be essentially unsecurable.

---

<sup>1</sup> Quinn (1993), pp. 109-33

But no matter how we navigate those options, the fact remains that, at present, most of our everyday moral judgments are unsecured. I am not yet prepared to conclusively deny the possibility of securing moral judgment, and I am open to possibility that the epistemic status of most moral judgments may change in the future, in which case any analysis I present here will simply be an epistemic interim solution. I suspect, though, that more epistemic work will someday show that moral judgments are essentially unsecurable.

We have historically been suspicious of unsecured judgments. But, I will argue, recent epistemic thinking has revealed that most or all human knowledge is actually unsecured, including such cognitive stalwarts as sensory perception and memory. We must, as a consequence, accept unsecured judgments or lose all rights to knowledge and even cognition in general. Any theory that accepts the unsecurability of these cognitive basics as epistemically valid entryways into knowledge will also, at least in principle, have room for unsecured moral judgments.

### **A brief introduction to entitlement theory**

At present, there are two standard responses to moral disagreement over unsecured moral judgments. The first, typical of moral nihilists like Gil Harman and J.L. Mackie, is to claim that the pervasiveness of unresolved moral disagreement, combined with their unsecured nature, shows that there are no genuine moral facts - that our moral intuitions are merely subjective phenomena. The second, typical of many others, is to



disregard moral disagreement entirely. Both responses are, I claim, epistemically naïve; both arise from on an essential disappointment with the unsecurability of moral judgment. Much of the problem comes from a historical lack of any satisfying epistemic account of moral judgment, especially unsecured moral judgment. A move to moral cognitivism, then, seemed to demand that we isolate ethics from epistemology. But recent developments in epistemology promise to fill in this gap. The most plausible modern epistemic theory, entitlement theory, is devoted to the analysis of the many forms of knowledge that have unsecured grounds. By applying the developments of entitlement theory to moral reasoning, I think we can carve a reasonable middle path between nihilism and dismissal.

Entitlement theory arises in opposition to the Cartesian approach to knowledge. To paint with a broad brush, the Cartesian approach demands that we secure our foundations - that we provide a proof or account for any and all of our beliefs. But this demand seems impossible to satisfy. If we begin by distrusting all our beliefs and faculties until given a reason to trust them, then we shall never trust them, for any possible supporting reasons could only arise from some already trusted belief or faculty.<sup>2</sup>

Let us consider the following principle; call it the Prior Justification Demand. The principle states: *in order to reasonably trust a given mental ability, we must have a prior account giving us a reason to think that ability is reliable.* It is something like this principle that's behind both the Cartesian approach to knowledge and justification, and the ensuing skepticisms that seem to plague all attempts to provide a complete Cartesian

---

<sup>2</sup> The clearest exposition of this thought occurs in Wright (2004), pp. 168-175.

justification of any piece of knowledge. This is easy to see in the perceptual realm. We have some perceptual experience - it seems to us that we see an apple and smell an apple. Suppose the Prior Justification Demand is correct. It seems like any account of the proper functioning of our perceptual faculties - our visual system, our olfactory system - will have to make some references to a scientific theory - to optics, chemistry, to neurobiology. But those scientific theories themselves depend on data gathered from the senses. A vicious circle threatens. If we demand a prior account showing us that our perceptual system works before we are willing to accept it, we'll never get one, because the very sciences that might vet our perceptual systems themselves depend on a basic trust in our perceptual systems.

This circularity is even more apparent when we shift our focus to our basic cognitive abilities, and why we trust them. If we demand a convincing *argument* for trusting our mental abilities before we start to trust them then we'll never get anywhere at all, because the very act of finding and evaluating arguments itself depends on some cognitive abilities. Tyler Burge, in "Content Preservation," argues convincingly that all reasoning depends on *trusting* some substantial cognitive abilities.<sup>3</sup> If we do the reasoning in our head, we're trusting our short-term memory. When an agent mentally rehearses the steps of the argument, she mentally focuses on a single step, and then stores the content of the step in short-term memory as she moves to the next step. If she doesn't trust her short-term memory, then she might write down the steps - but then she's simply shifted her dependencies to her abilities to write, see, and read. Due to the narrowness of

---

<sup>3</sup> Burge (1993)

human consciousness - the fact that we can only really hold a few words at the forefront of our consciousness at once - we must rely on some storage facility for any complex reasoning.

Once we see that any act of reasoning depends trusting some mental abilities, we can see that the problem lies with the Prior Justification Demand itself. If we cannot trust any mental ability unless we have an argument telling us to do so, and if finding and evaluating arguments depends on mental abilities, we will never be able to trust our mental abilities. It follows from the Prior Justification Demand that we cannot reasonably trust any belief. The proper response, suggests entitlement theory, is not to give up on the possibility of knowledge or even reasonable belief, but to reject the Prior Justification Demand. We must flip the burden of proof. Instead of adopting a default stance of distrust towards all cognitive abilities, we ought, instead, trust our cognitive abilities until we discover reasons why we should not.<sup>4</sup> The central insight of entitlement theory is that cognition in all domains involves initial self-trust, adopted without proof.

Thus, suggests entitlement theory, we are entitled to start cognitive life by trusting our mental abilities and thus trusting that things are as they seem to us, even if we cannot provide a proof that our senses and mental abilities are, in fact, reliable.<sup>5</sup> This trust is, however, only tentative; it is vulnerable to future defeat. Unlike the intuitionism of an earlier era, entitlement theory doesn't purport to offer us unshakeable foundations for knowledge. Entitlement theory instead claims that our cognitive faculties are fallible, but

---

<sup>4</sup> The reasoning here is not simply practical. The Cartesian approach to knowledge depends on holding the principle P, that every belief, to be reasonably held, be deducible from another reasonable belief. If P is true, then it self-applies, and this shows that it is not reasonable to hold P. Therefore, we ought to hold P false, and drop the universal demands for grounds.

<sup>5</sup> Pryor (2000)

admits that trusting them is the best we can do. Cartesianism's epistemic miserliness leads to skepticism; entitlement theory substitutes, not an epistemic free-for-all, but an attitude of qualified epistemic generosity. This generosity entails that the default stance towards our own mental abilities is one of tentative acceptance. In contemporary epistemic parlance, this stance is called *defeasible entitlement*. We're entitled in assuming our mental abilities work, but, since we don't have an account demonstrating that those mental abilities are, in fact, reliable, this entitlement is vulnerable to future evidence that shows that it isn't so reliable after all.

Entitlement theory strikes me as the most plausible epistemic theory presently available. I claim that entitlement theory extends to unsecured moral judgments. I will expand upon my description of entitlement theory, and argue for its extension to moral cognition, in Chapter 2. Loosely, entitlement theories tell us that our default stance towards our cognitive abilities should be one of acceptance. We should accept that things are as they seem, in both simple perceptual cases and complex cases of rough judgments. Insofar as our unsecured moral judgments make claims about what is and isn't the case, then they fall under the license of entitlement theory. This is not a special exception for unsecured moral judgments. Rather, entitlement theory shows us that the initially unsecured nature of moral judgments is not different in kind from initially unsecured empirical judgments.

Entitlement theory will give us the means to describe a reasonable middle ground between the nihilists and the dogmatists. I will argue that once we apply entitlement theory to unsecured moral judgment, we will see that we cannot regard ourselves as

morally self-sufficient. Our moral judgments are defeasible, and vulnerable to defeat from social sources. I will present my argument for this claim in Chapter 2.

My picture of our moral epistemic condition is mildly pessimistic, but not nihilist. I do not think that we have no reason to trust ourselves; I think, rather, that we have plenty of reason for initial moral self-trust, but we also encounter plenty of evidence to reduce that self-trust. The result is, at present, a painfully unstable situation. I find myself compelled both by the evidence of my moral intuitions, and compelled by epistemic considerations to the contrary. I cannot give up on either; the only remaining path is a tormented sort of half-certainty.

### **Part III. Moral humility and its conditions**

Let me now explain my thesis in greater detail. I claim that we are not morally self-sufficient; that the mere existence of disagreement, in and of itself, can and should reduce our confidence in our own judgments, when those judgments are unsecured. I'm arguing for the epistemic and moral acceptability of *social pathways to moral knowledge* - that, to the extent that we are cognitively fallible beings committed to cognitive moral beliefs, we can and should use social methodologies as sources of information about our own cognitive reliability. My claim is not that the moral domain is epistemically unique; rather, I am claiming that moral judgment is of a piece with other domains of knowledge, and subject to the same methodologies of self-evaluation and self-correction.

Disagreement with peers is a form of disconfirmation and thus vital data for evaluating one's own reliability, in moral disagreement as well as in empirical disagreement.

My central claim is that the mere existence of certain types of moral disagreements ought bring us to reduce our confidence in those beliefs over which there is disagreement. I do not claim that all moral disagreement ought to reduce confidence, only that moral disagreement is the right sort of thing to cause us to reduce confidence, and it ought to do so when the conditions are right. These conditions are that the disagreements are long-standing, unresolved, and with peers that we respect.

I do not claim that moral disagreement may bring us to *suspend* some of our beliefs. I suspect, but have not yet demonstrated to my satisfaction, that there may be an upper bound to the epistemic effects of disagreement. It strikes me as likely that moral disagreement will, at most, raise only reasonable doubt towards our own belief, and never bring us to actually suspending a belief outright. But I haven't settled this matter, so I limit myself to showing that disagreement will bring at the very least doubt and suspicion towards our own beliefs.

### **Moral peers and best cases**

To establish that moral disagreements are the right sort of thing to have epistemic effects, I will focus on what I take to be the best-case scenario for epistemic effects: long-standing, intractable disagreements with our moral peers. By "moral peers", I mean those

we take to be approximately at least as morally sensitive, reliable, rational and trustworthy as ourselves.

This is, I believe, a significant break from the contemporary discussion of moral disagreement, which largely focuses on the most radical forms of moral disagreement - the saints versus the psychopaths, so to speak. But, while these examples are quite striking, they aren't actually the best cases for epistemic effects. When a kind, good, and charitable agent is locked in a disagreement with a serial-killing psychopath, the saint has very little reason to give any epistemic credence to the psychopath whatsoever. They share no values; the psychopath seems to have no sensitivity towards or understanding of the moral domain. In straightforward empirical domains, the most epistemically compelling cases of disagreement are those that occur between agents that have every reason to trust one another, including, by and large, agreement on most issues. A NASA physicist has very little reason to reduce her confidence if she discovers she has extensive disagreements with an aluminum-foil-hat wearing cultist ranting about alien mind-control rays, but has very good reason to reduce confidence in her beliefs if she discovers some disagreements with fellow astrophysicists, professors of physics, and NASA engineers, with whom she shares a similar training and background beliefs. I am not claiming that there can be no epistemic effect from cases of extreme, foundational disagreement - but rather, that those types of cases are edge cases, and not the best examples to show the in-principle epistemic importance of moral disagreement.

The best cases will also be cases of intractable disagreement. When I first discover the existence of a disagreement, there are many possible explanations. Suppose

I'm entirely certain that there's no good tea in Los Angeles. My colleague mentions, as she passes down the hall, that she just had some good tea down the street. At the first moment, there are many plausible explanations for the disagreement. Perhaps she meant something entirely different by "good" - perhaps I mean "delicious" and she means "cheap and fast." Perhaps she doesn't have the vast experience and expertise that I have with tea. Perhaps she cares about things I simply don't care about - presentation and service and the presence of finger-sandwiches. There are many factors that might explain away the disagreement. But the more we talk it through, the more I discover that we do mean the same thing, that she does have as much experience as I do and cares about the same things I do, the more the disagreement becomes epistemically threatening. The more I try to explain away the disagreement and fail - the more intractable the disagreement turns out to be - the more reason I have to doubt. The doubt will be diffuse and spread over many possible explanations. I have reason to doubt the reliability of my own taste, the reliability of her taste, even to doubt the objectivity of the whole realm of taste in tea. But all of these possibilities constitute markers against my original, all-fired certainty in the poor quality of tea in Los Angeles.

I suggest, then, that we take up as the best cases for epistemic importance cases of long-standing, intractable moral disagreement between peers. Such cases, though they don't often appear in the contemporary discussion, are actually quite familiar cases. They are the cases that intuitively give the best reasons for worrying about the reliability of our moral judgment. The best reason to doubt my moral judgment isn't when I discover that the Unabomber has a different moral code than me; it's when I discover that, say, my



advisor, the profoundly thoughtful and compassionate Kantian ethicist with whom I largely agree on most moral issues, has a strikingly different attitude than me towards what's permissible in warfare.

Let me present three illustrative cases. Imagine, first, that a species of endangered owls have been discovered in a logging forest. We now have a dilemma: ought we cease logging for the sake of the owls, thus cutting off the main income source for the small town nearby, or ought we allow the logging to continue for the sake of the logging community? Imagine that two groups of very liberal, progressive activists come into conflict over this issue, and, despite lengthy discussions between the activists, remain firmly entrenched on their respective sides. Notice that activists on either side of the debate are obviously thoughtful, morally committed, sensitive people, and likely accept the other side as similarly morally interested. In fact, both sides of the debate surely acknowledge the values and commitments of the other - the environmentalists surely think that jobs and families are important, and the community advocates surely think that the preservation of endangered species is important. They only disagree about which of these moral considerations is more pressing in this case.

Second, a case familiar to many: my friend David and I are both graduate students in ethics. We've both thought a lot about ethical issues, and we both try to be good. David's a vegetarian, and I'm not. Again, obviously both sides are morally committed and recognize each other as such. This case is particularly interesting because it's not simply a case of bare intuitions are clashing. David isn't immediately struck by the awfulness of eating meat; in fact, he ate meat for most of his life, and still craves it. He has reasons that

brought him to, with much tribulation and backsliding, slowly give up meat. We both agree that animal suffering is important, and I try to avoid, for example, eating veal raised in cruel conditions. But he thinks that the raising and harvesting of animals for our mere gustatory pleasure is wrong - that animal shouldn't be used in that way - and I simply do not. I have heard his arguments, and am simply not convinced by them.

Third, a case that's particularly important to my own biography: I'm Vietnamese. I see the fact that something is traditional to Vietnamese people as no particular reason to do it - especially, for example, the directive to marry in-race. My father, who is a very good, kind, thoughtful person, thinks it's obvious that tradition is important, that it's major source of value and direction. Furthermore, he thinks that it is traditional, and therefore very important, for Vietnamese people to marry in-race. This sort of case strikes me as a conflict of direct intuitions about fairly abstract values. My father simply finds utterly obvious that tradition is a major source of reason and value. I cannot even imagine my way into it - it seems strange and alien, and I have to, at best, take his word for it that it's important to him. Not only do I not feel the value of tradition or find it compelling, but I can't even imagine what it would be like to find it compelling. Some values we don't feel, but can imagine what it would be like to feel. I do not care about neatness in my life, but I've had the occasional glimmer of something feeling better when I organize it to imaginatively expand it in my head, and imagine what it would be like to find neatness deeply important in one's life. But tradition is blank for me - I simply see nothing there, and continue to despite long discussion. When one friend of mine goes through tortures

because they are Jewish and they are in love with somebody who is not, and they cannot break with their traditions to marry them, I am just flabbergasted; I simply do not *see*.

These are all cases where there are disagreements between people who are or might be familiar with each other, and all cases where the people involved have talked through the disagreement and found it to be intractable. Notice that the cases I've described fall into two categories. In cases like the argument of tradition, one party of the disagreement claims that there is a certain value, and the other claims that there isn't. Such disagreements are quite vivid, and are the sort usually discussed in the disagreement literature. And while there are some disagreements over the very existence of a value, most disagreements are more like the environmentalist case: they are cases where there are two competing values, which both parties subscribe to. In the environmentalist case, we needed to ascertain which value trumps which other value. But there are many situations where there is no binary decision - where the interplay of values leads to a decision about *how much* and not *whether to*. For example, I may agree with my girlfriend that it is rather bad that I forgot her birthday, and that she is justified in being angry with me, but we may disagree rather strongly about *how bad* it was and *how much* anger is justified.

Focusing on these realistic situations of peer disagreement will, I think, do much to make my conclusion more palatable.

## **Consequences, practical and internal**

If I can establish that there is a class of epistemically active moral disagreement, there will be significant consequences. I take it as an empirical fact of most of our lives that there are a very large number of such disagreements over a wide array of topics. Even though an agent may rate only a small minority of others as their moral peers, for most agents there will still be a great number of moral peers. Furthermore, even if we imposed a very strict standard for peerhood - agreement across a majority of judgments, for example - we'll likely still be able to discover disagreement over a significant percentage of our moral beliefs, between the many peers. I disagree with one of my peers about vegetarianism, with another about the importance of maintaining ethnic traditions, and with another about the permissibility of military intervention.

If it's true that we can find such disagreements over most of our moral judgments, then we will have arrived at a qualified form of moral skepticism - not the radical skepticism of the nihilists, but a constrained skepticism, a skepticism of reasonable doubt and moderate suspicion. Notice that this constrained skepticism is not born of the mere possibility of failure; it is born of substantive empirical evidence raising the probability of malfunction in our own process of judgment. The mere possibility of failure opens the door to doubt; it is the actual existence of disagreement that walks through that door.

Since this skepticism is contingent on acquiring certain types of empirical evidence, the amount and degree of suspicion engendered by disagreement will vary from one person to another. One agent may rate almost everybody their moral peer, another may, for very good reason, consider very few to be their moral peers. A very

cosmopolitan person may run across a larger number of disagreements than somebody who has never left their small home town.<sup>6</sup>

There are practical consequences to this conclusion. Even if we discover that disagreement can never bring outright suspension of belief, and the socially acquired moral humility is limited to self-suspicion and self-doubt, there will still be practical consequences. The reasonableness of actions is complexly related to our degree of confidence in our justifying beliefs. Some actions can be justified by the merest hint of a reason; other actions require very high degrees of confidence. Even if I think my ability to predict when the bus is coming is only slightly reliable, I can still act on my belief. On the other hand, highly consequential, irreversible actions - like taking another life or calling for a nuclear strike - seem to require a very high degree of confidence. Thus, the presence of even a mild doubt may bring us to refrain from action. For the same reason, we require belief beyond a reasonable doubt in, for example, calling for the death penalty. Even if disagreement can bring only doubt and not the suspension of belief, it could sometimes give us a reason to refrain from action. The extent of practical effects will depend on the precise principles relating actions to certainty; this is an important area for future research.

I take the conclusion of the arguments presented in this dissertation to be quite narrow. I am only discussing best-case scenarios: intractable disagreements with clear peers. It seems clear that the impact of moral disagreement is much broader than these best-case type disagreements, but I cannot yet say how much broader. There are many

---

<sup>6</sup> Whether or not there is a further rational imperative to seek out more interlocutors and disagreement is an open question, though I suspect that there is.

further questions that must be settled to ascertain the range of impact of moral disagreement. What, for example, are the outer bounds of peerhood? Can we reasonably accept another as a moral peer on very general considerations (attention, care, thoughtfulness) even though we disagree with them on many particular judgments? Another vital topic for future inquiry is the infectiousness of doubt. I only take myself to establish the relatively weak claim that a particular disagreement over a judgment creates suspicion over that particular judgment. I do not think this is the end of the matter; there must be some pathway for belief-specific doubts to infect outwards. I can only be utterly certain of my ability to remember the calculus I learned ten years ago for so long when I start discovering calculus mistakes; once I make enough specific mistakes, I should start doubting my ability to do calculus in general. Similarly, if I have enough failures of my memory for people's faces, I should probably start worrying about any similar instance. Since the best explanation for a judgment's being incorrect is that the underlying faculty or ability is unreliable, and this unreliability should spread throughout the faculty or ability. However, it is at present difficult to say exactly how the doubt should spread, especially in those cases where we do not have a clear account of what precisely faculties or abilities underlie our judgments. Though it does seem likely that doubt will infect outwards from the particular judgment, the precise degree and extent of infectiousness needs further elaboration.

My constant deferrals to future research may strike some as punting on the core issues. I think, rather, that the very reason that topics of moral peerhood, the infectiousness of moral uncertainty, and action in situations of moral uncertainty, are

underexplored is precisely because ethical philosophy has thus far treated moral disagreement as unimportant. By arguing for a proof-of-concept - by showing that moral disagreement is the right sort of thing to inspire changes in confidence - I hope to show that topics concerning social pathways to moral knowledge are relevant and vital. One of my primary goals is to show that contemporary ethical theory has neglected of the most important areas of study: how we are to behave from a position of profound, and potentially irresolvable, moral uncertainty. We cannot, in good conscience, presently treat ourselves as perfectly morally reliable, yet we are perpetually called upon to act. How unsure ought we be; how ought this uncertainty infect our actions? This is a vital direction for future research.

But even if I cannot yet map the precise bounds of the practical consequences, I take it that my narrow conclusion will show the crucial internal consequence: the attitude of moral humility. Even if we cannot say precisely who is our peer and even if we cannot say precisely how our doubt will infect outwards, we can say this: there are some people who are clearly our peers, who have the power to drag our moral lives into the realm of suspicion. We cannot proceed with unfettered confidence; at best, we can proceed with a queasy half-certainty, aware that we have plenty of reason to doubt our beliefs, and unable to show where precisely those doubts end. Disagreement raises doubt, and until we can resolve those disagreements, we must live much of our ethical life inside the shadow of doubt.

## **A brief summary of what is to come**

In the coming chapters, I will present the following arguments.

In Chapter 2, I will argue that our entitlement to self-trust does in fact extend to the moral domain. Though we are reasonable to trust our moral judgments, that judgment is defeasible. I will then argue that moral disagreement is a form of partial defeat of our entitlement to trust our own moral judgments. My argument will be that if there are reliable moral sources with whom we disagree, then we will have reason to doubt. Insofar as we take moral judgments to be of objective truths, then disagreement between two reliable sources is a reason to distrust both reliable sources.

The significant question is, then, whether or not there are really such things as reliable moral sources. It seems intuitive that there are. There are people whom we turn to for advice, whom we trust for judgments on the moral character of others, who we trust to tell us when we have gone too far, acted too much in anger. More importantly, it follows from the start in entitlement theory that there are moral peers. Entitlement theory permits us to trust our moral intuitions. The relevant moral intuitions present themselves as of objective content. In trusting our moral intuitions, we are implicitly taking our moral abilities to be reliable. Now we turn our gaze to other people. Ought we trust them? The answer is: given what entitlement theory says, yes, we should trust some of them, because they are so similar to us.

In the situation of entitlement, we trust our moral judgment to be reliable, but we have no account of which particular cognitive abilities are responsible for that reliability. Other people are cognitively more like us than not: they seem to have similar abilities,



similar physical apparatus, similar biological and educational backgrounds. For the prime candidates for moral peerhood, we have every reason to rate them similarly reliable, and little reason to rate them unreliable. We must rely on these general features of rationality, because with an entitlement start, we have no particular account to help us pick out which cognitive features are more or less relevant to the assessment of moral reliability. Thus the entitlement to trust myself in some domain extends to trusting others, based on *empirical evidence that they are creatures substantially like me*. In short, reflection on our entitlements forces each of us to admit that we are but one cognitive agent among many. An agent has no a priori reason to think that she has fundamentally better access to the moral facts than another agent.

From the entitlement start to moral self-trust, we must admit that our self-trust is vulnerable. Moral self-sufficiency cannot stand on epistemic grounds, because we are each only a single, finite, fallible truth-seeker among many. The mere fact that an interlocutor disagrees is enough to bring doubt - I don't have to understand his argument, or have been brought around by his reasoning. Since our moral peers are reliable, the mere fact that they believe the opposite from what we do should, in and of itself, worry us.

The above constitutes a summary of my positive argument. In Chapter 3 and 4, I will consider what I take to be the two primary objections to my position. In Chapter 3, I will consider the claim that disagreement never has epistemic weight in and of itself, on epistemic grounds. In Chapter 4, I will consider the claim that, though disagreement may give us good epistemic grounds to reduce our confidence, these epistemic grounds are

trumped by moral considerations, which makes it imperative to ignore social pathways to moral knowledge.

## Chapter 2

### The Epistemology of Moral Humility

The present range of moral disagreements should be far more threatening to our confidence than the empirical disagreements. Our empirical judgments are far better corroborated and well-explained. Our moral judgments, on the other hand, are utterly lacking any stable, complete, universally accepted explanatory theory. Furthermore, moral disagreements are more pervasive and more varied than empirical disagreements; we have disagreements with peers on a wide range of basic judgments. Thus, moral disagreement ought to have a significantly detrimental effect on our moral self-confidence. It seems as though moral disagreement ought to reduce our moral self-confidence.

However, I seem to be in the minority in advocating *moral humility* through disagreement. Most believe that moral disagreement doesn't matter. Instead, the most common position seems to be that agents should be *morally self-sufficient*; that agents ought not rely on the beliefs of others in establishing their own moral judgments, nor should they let the mere fact that others agree or disagree change their degree of confidence in those judgments. The position of moral self-sufficiency therefore depends on establishing that there is crucial disanalogy between empirical reasoning and moral reasoning.

In this chapter, I will argue that the epistemic grounding for moral reasoning and empirical reasoning is essentially similar, and that no such radical disanalogy holds.

Though I will grant that there may be reasons to block our forming new beliefs through testimony, I do not think there are reasons to block the loss of self-confidence through disagreement. Given the best currently available epistemic theory, we must treat our own reasoning as fallible, including our moral reasoning. Since there are others who have as much in-principle access to the moral domain as we do, and since all evidence indicates that their reasoning is of the same kind, and of approximately the same reliability, as our own, disagreement with them ought to have some effect on our all-things-considered moral beliefs.

My argument consists of two stages. First, I will argue that a contemporary epistemic theory, entitlement theory, gives us the most plausible moral epistemology. Second, I will argue that, given entitlement theory, we must treat others as moral peers, and, consequently, we must pay attention to disagreement. I take this chapter to present my entire positive argument for moral humility and against moral self-sufficiency; the subsequent chapters will take up the major objections.

## **Part I. Entitlement Theory and Moral Judgment**

The position of moral self-sufficiency depends for its plausibility on a naive moral epistemology. The position has survived, it seems to me, because much of modern ethics occurs in an epistemic vacuum. The divorcing of ethics and epistemology is quite understandable, for no straightforwardly foundationalist epistemic theory has been able to adequately capture our basic intuitions about how moral reasoning should go. Traditional

epistemic theories have usually demanded explicit proof; our moral lives are full of moral self-confidence, but precious little in the way of *proof*. Faced with a choice between giving up on our moral confidence entirely or cutting moral reasoning loose from epistemic considerations, most of us, understandably, choose to turn away from worrying about the epistemology. But recent developments in epistemology will do much to repair this situation. Entitlement theory, which I believe to be the most plausible contemporary epistemic theory, will neatly capture many of our basic intuitions about how moral reasoning should go; it will show how some amount of moral self-confidence is reasonable even in the absence of proof. But entitlement theory only grants us tentative self-confidence; it is incompatible with moral self-sufficiency.

First, let me specify the domain of discussion. My arguments concern only our confidence in *unsecured moral judgment* - moral judgments for which we have no explicit proof. This includes, as I argued in Chapter 1, both raw moral intuitions and more complex forms of judgment. I also confine my discussion those judgments that present themselves as making claims about objective affairs. Though not all moral judgments present themselves as objective in this way, a claim to objectivity is prerequisite for the existence of moral disagreement. The presence of disagreement presupposes a single object over which there is disagreement.<sup>7</sup>

I am not yet claiming that our moral judgments *are* of objective affairs; only that they present themselves as being so. I am also not claiming that purported objectivity is a

---

<sup>7</sup> Of course, Tom and Quentin could have a disagreement about whether Quentin actually likes pomegranate or is just pretending, but this is a psychological disagreement, not a moral disagreement. I take it that the existence of distinctively moral disagreement presupposes, at the very least, that the involved parties are committed to the existence of some objective moral facts in some minimal sense.

necessary part of any moral judgment or intuition; rather, I claim it as a matter of empirical fact about the phenomenology of many people's moral experience. When I see a person beating their dog, it seems to me wrong, and objectively wrong. When I first meet somebody that doesn't share my judgment that beating dogs is wrong, I am *shocked*. My naïve moral phenomenology is of rightness and wrongness as plain, public, objective affairs.

### **Introducing entitlement theory**

The problem with providing an epistemic grounding for moral reasoning is that so much of moral reasoning is unsecured, and, historically, most epistemic theories demand security. But entitlement theory is distinctive in this regard; entitlement theory is an epistemic theory with thoughtfully lax standards of security. Entitlement theory arises from the observation that virtually all knowledge requires trust in unsecured judgments. We can, for example, assess the reliability of our eyes only if we possess a science of optics, but such a science could only come through trusting and using our eyes. The only way to possess any form of science is if we begin with an attitude of trust - an unsecured trust - towards many of our empirical faculties and cognitive abilities. It might have seemed that moral knowledge was uniquely unsecured, but, if we accept entitlement theory's analysis, moral knowledge turns to be not so peculiar after all.

Let's begin with a more detailed look at entitlement theory. Entitlement theory is the body of epistemic theories that claim that knowledge and reasonable belief must

begin with some degree of cognitive self-trust, even in the absence of secured foundations. The best overview of entitlement theory comes from Crispin Wright, in his, "Warrant for Nothing (and Foundations for Free)?" Wright dissects older epistemic theories by showing how their basic methodology must necessarily lead to skepticism.

Call a proposition a *cornerstone* for a given region of thought just in case it would follow from a lack of warrant for it that one could not rationally claim warrant for *any* belief in the region. The best - most challenging, most interesting - skeptical paradoxes work in two steps: by (i) making a case that a certain proposition (or restricted type of proposition) that we characteristically accept is indeed such a cornerstone for a much wider class of beliefs, and then (ii) arguing that we have no warrant for it.<sup>8</sup>

Typically, skeptics argue for (ii) by claiming that the cornerstone depends for its justifications on other claims from the very domain for which it is the cornerstone.

...There is a vicious circle: it is only if I can get a warrant for a specific proposition about it that I can acquire a warranted belief that there is a material world, yet it is only if the latter is already warranted and part of my collateral information that I can draw on my experience to provide warrant for specific beliefs about it.<sup>9</sup>

This, says Wright, is the formal structure of virtually all skeptical arguments, including skepticism towards other minds, the external world, and the existence of past events.

Take, for instance, skepticism towards the external world. Typically, an external world skeptic begins by noting that our knowledge of the external world depends on the thesis that our senses are reliable. The skeptic simply needs to show that the thesis of the reliability of our senses depends for its justification on some facts about the external world - for example, that there is no omnipotent evil deceiver, or that the biological apparatus of our sense organs is well-functioning. But both of these requisite facts are

---

<sup>8</sup> Wright (2004), p 168

<sup>9</sup> Wright (2004), p 171

themselves facts about the external world. Insofar as our having *any* warranted beliefs about the external world depends on having *some particular* warranted belief about the external world, then we fall inescapably into the vicious circle.

The problem is not confined to skepticism towards the external world. Tyler Burge argues that this form of problem extends to virtually any form of cognition. In "Content Preservation," Burge argues that any form of reasoning with multiple steps relies on the substantive use of fallible cognitive faculties. We are either holding the steps in our head, in which case we are using short-term memory, or we are writing down the steps, in which case we are using our ability to see and read. Burge is pointing out that *reasoning* of any non-trivially complex form depends on the well-functioning of an array of fallible cognitive faculties and abilities. But if we demand that we possess a justification for trusting these cognitive abilities *before* we begin using them, then we fall again into the vicious circle, for all possible justifications involve the use of those very abilities. Thus, argues Burge, all cognitive life must begin with an entitlement to trust our basic cognitive abilities without justification.

I accept Burge's argument. There seems to be a vast array of substantive, fallible cognitive abilities employed in any non-trivial reasoning, of which short-term memory is but one clear example. The argument concerning short-term memory is a particularly useful one, because the use of short-term memory is clear and obviously pervasive. Burge's analysis shows that we need entitlements not only for empirical knowledge, but for any form of extended reasoning whatsoever. Almost all reasoning depends on the well-functioning of some cognitive abilities. If we demand that any agent possess a



justification for trusting in the well-functioning of that ability before they may reasonably trust the ability, then we will never gain such a justification. In the empirical realm, we use the empirical sciences to vet the reliability of our senses, but we can only build those sciences using our senses. For any complex cognition we must rely on short-term memory, but any justification we might provide for the reliability of our short-term memory seems to require some form of complex cognition. The demand for prior justification for fundamental cognitive abilities leads inevitably a vicious circle, and thus to skepticism.

Entitlement theory treats this analysis as a *reductio* of a stringent demand for prior justification; for knowledge to be possible, we must begin without justifications for all our starting points. James Pryor considers his version entitlement theory to be a development of G. E. Moore's discussion of common sense. Moore, famously, claimed that he could know it to be true that the external world existed; he could prove this because he knew that he had his two hands in front of him. But Moore famously refused to offer any proof of that latter piece of knowledge. Writes Pryor,

...Moore doesn't seem ready or able to offer *any considerations at all* in favor of the claim that he has a hand - even defeasible, ampliative considerations - without begging the question against a skeptic who refuses at this stage of the dialectic to grant the existence of the external world. This is why Moore's "proof" strikes us as so unsatisfactory: he hasn't offered any non-question-begging reasons to believe his premise. Yet Moore claims he can know these premises to be true. He can know them to be true, though he has no non-question-begging arguments to offer in their support.<sup>10</sup>

Moore is being reasonable only if his refusal is reasonable, and this refusal is reasonable only if Moore doesn't owe us considerations in favor of believing that his hands are in

---

<sup>10</sup> Pryor (2004), p 518

front of him. Entitlement theory offers us an explanation of why the refusal is reasonable: it is because the burden of proof is not on Moore to provide a proof of every claim. Rather, the burden of proof must lie on the skeptic's side, to give Moore a reason why he ought to take up the question.

Entitlement theory suggests that, instead of beginning our cognitive lives by distrusting our cognitive abilities until given a reason to trust, we initially trust our cognitive abilities until given a reason not to. I believe that entitlement theory is the best epistemic theory in the running. I will presume that the basic insight of entitlement theory is correct: that we must begin our cognitive lives with an initial attitude of epistemic generosity rather than an attitude of epistemic miserliness. I will provide no further direct arguments for entitlement theory; such arguments can be found elsewhere.<sup>11</sup> However, though I take the general approach of entitlement theory to be correct, I don't take any of the presently available theories to be entirely satisfying.<sup>12</sup> I will attempt to show is that the general approach of entitlement theory is quite plausible as an epistemic basis for moral judgment, and that this is enough to show that moral self-sufficiency is wrong.

### **Entitlement and moral judgment**

I would like to show that entitlement theory applies to unsecured moral judgment. Since I'm not committing myself to a specific entitlement theory, I won't be able to argue this definitively. I'll do the best I can under the circumstances: I will introduce a number

---

<sup>11</sup> Burge (1993) and Pryor (2004)

<sup>12</sup> Wright (2004) provides an overview of what he takes to be the primary candidates in the running for entitlement theories, and shows why none of them are satisfying. I am in agreement with Wright's conclusions - that some form of entitlement theory has to be true, but that nobody has yet formulated a satisfying specific theory.

of representative entitlement theories currently in play, and show that each of those easily extends to moral judgment. Then I'll give a set of considerations why it seems likely that any reasonable entitlement theory will so extend.

It will be easiest to see *why* entitlement theory ought to extend to moral judgments if we look at some of the earliest, most intuitively straightforward versions of entitlement theory. According to Wright, one of the earliest versions of entitlement theory was developed by Hans Reichenbach. Reichenbach proposed what Wright terms a "strategic entitlement." According to Reichenbach, an agent is strategically entitled to believe *p* if believing *p* is part of one's *dominant strategy* - in other words, if one's best practical strategy is to act as if one believed *p*. Imagine, says Wright, Robinson Crusoe stranded on the island, confronted with a new fruit. He is dying of starvation. Crusoe has no reason to think that the fruit is poisonous rather than edible, or vice versa. He is, however, warranted in acting as if the fruit is edible and eating it, since he would be no worse off if the presumption were wrong than if he refused the presumption, and significantly better off if the presumption were right. Since starvation will kill him just as surely as poison, epistemic conservatism here has no rewards. Pragmatism warrants his acting as if he believed that the fruit weren't poisonous.

Trusting our senses is also a dominant strategy, says Reichenbach. Our senses are our sole access points to the empirical domain. If we don't trust our senses, we have no access to empirical facts at all. If we trust our senses and they are unreliable, then we are no worse off than if we didn't trust them. Since the worst-case scenario of trusting our

senses is no worse than what will happen if we refuse to trust and the best-case scenario is considerably better, we are warranted in acting as if our senses were reliable.<sup>13</sup>

The argument works on any set of abilities which are purportedly the sole access to some domain. Reichenbach's legitimization of dominant strategies is the dressed-up version of an old response to skepticism, often made by non-philosophers: "Well, what other options do we have besides trusting our senses?" This pragmatic argument, of course, will have many detractors. Wright's chief objection is that the pragmatic argument can make reasonable *acting as if* it were true that the fruit was not poisonous, but couldn't make reasonable the belief that the fruit was not poisonous. As Wright worries, it is a pragmatic entitlement, but not an evidential entitlement.<sup>14</sup> But if one finds Reichenbachean pragmatic entitlements acceptable, note that they easily extend to moral judgments. In the absence of a developed moral theory, our loose moral judgments are the only access we have to the domain of moral facts. If we trust and our moral judgments are incorrect of the facts, our actions will be as good as random as to what we actually ought do - just as if we took our moral judgments to be unreliable. In fact, if we trust our moral judgments and it turns out that we are deeply wrong, and there are no moral facts whatsoever, then we still have taken no morally *wrong* action. Thus, trusting our loose moral judgments is the dominant strategy, and so warranted.

The other progenitor of entitlement theory is Wittgenstein. Wright reads Wittgenstein in the following way:

---

<sup>13</sup> Wright (2004), p 178-184

<sup>14</sup> Wright (2004), p 184-188

To take it that one has acquired a justification for a particular proposition by the appropriate exercise of certain appropriate cognitive capacities—perception, introspection, memory, or intellection, for instance—always involves various kinds of presupposition. These presuppositions will include the proper functioning of the relevant cognitive capacities, the suitability of the occasion and circumstances for their effective function, and indeed the integrity of the very concepts involved in the formulation of the issue in question. I take Wittgenstein’s point in these admittedly not unequivocal passages to be that this is essential: one cannot but take certain such things for granted...

That is not to deny that, if one chose, one could investigate (at least some of) the presuppositions involved in a particular case. I might go and have my eyesight checked, for example. But the point is that in proceeding to such an investigation, one would then be forced to make further presuppositions of the same general kinds (for instance, that my eyes are functioning properly now, when I read the oculist’s report, perhaps with my new glasses on.) Wherever I get in position to claim justification for a proposition, I do so courtesy of specific presuppositions—about my own powers, and the prevailing circumstances, and my understanding of the issues involved—for which I will have no specific, earned evidence. This is a necessary truth. I may, in any particular case, set about gathering such evidence in turn—and that investigation may go badly, defeating the presuppositions that I originally made. But whether it does or doesn’t go badly, it will have its own so far unfounded presuppositions. Again: whenever claimable cognitive achievement takes place, it does so in a context of specific presuppositions which are not themselves an expression of any cognitive achievement to date.<sup>15</sup>

Wright suggests that Wittgenstein can be read as proposing an entitlement theory, which Wright calls the *entitlement of cognitive project*. One might, admits Wright, take these insights about the ineradicability of presuppositions to fuel a new sort of skeptical paradox. Better, says Wright, to swallow the bitter pill and keep going.

If there is no such thing as a process of warrant acquisition for each of whose specific presuppositions warrant has already been earned, it should not be reckoned to be part of the proper concept of an acquired warrant that it somehow aspire to this—incoherent—ideal. Rather, we should view each and every cognitive project as irreducibly involving elements of adventure—I have, as it were, to take a risk on the reliability of my senses, the conduciveness of the circumstances, etc., much as I take a risk on the continuing reliability of the steering, and the stability of the road surface every time I ride my bicycle. For as soon as I grant that I ought—ideally—to check the presuppositions of a project, even in a context in which there is no particular reason for concern about them, then I should agree *pari passu* that I ought in turn to check the presuppositions of the check—which is one more project after all—and so on indefinitely, unless at some point I can foresee arriving at presuppositions all of which are somehow safer than those of the initial project. If not, then there will be no principled stopping point to the process of checking: the quest for security will be endless, and therefore useless. And if that is the situation, then the right response—the reply will continue—is not to conclude that the acquisition of genuine warrant is impossible, but rather to insist that it does not require this elusive kind of security. Rather, warrant is acquired whenever investigation is undertaken in a fully responsible manner, and what the paradox shows is that full epistemic responsibility cannot, *per impossibile*, involve an investigation of every presupposition

---

<sup>15</sup> Wright (2004), p 189

whose falsity would defeat the claim to have acquired a warrant.<sup>16</sup>

I quote these passages at length partially because I think they are some of the most clearly articulated defenses of entitlement theory on record. Wright emphasizes that the complete ineradicability of presuppositions. One cannot but take certain things for granted in any cognitive project, including skeptical projects. This gives us a ready weapon to fend off the skeptic: her position is unstable. In order for a skeptic to convince a non-skeptic, the skeptic will have to present skeptical arguments; but such arguments themselves depend on substantive assumptions about cognitive functionality. Abstract reasoning, including the reasoning of the skeptic, does not get a pass from any proposed criteria for acceptable reasoning. Issuing challenges of philosophical skepticism is itself a cognitive project, which itself depends on presuppositions; excessively stringent prerequisites for knowledge cannot pass their own muster.

What Wittgenstein is saying, says Wright, is that one might have a reason to seek the justification for the presuppositions of one cognitive project as part of another cognitive project, but that undertaking another cognitive project will involve taking up a different set of presuppositions. This may be a good idea in some circumstances - for example, if I am engaged in the project of learning history in a class, I take on the presuppositions that my teacher and our history textbook are correct. But if I come to worry about these two sources, I can then challenge these presuppositions and seek justification - for example, by checking their professional credentials. But in doing so, I am engaging in a new cognitive project, which involves taking on presuppositions about

---

<sup>16</sup> Wright (2004), pp 190-1

the trustworthiness of the credentialing institution. It is possible, then, to justify the presuppositions of one cognitive project by taking on a different set of presuppositions, but it is impossible to do without presuppositions entirely. The right response to this insight, says Wright, is not to abandon all hope of knowledge or reasonable belief, but to realize that the "elusive security" of the eradication of all potentially flawed presuppositions is impossible, and so not a reasonable criterion for the acquisition of warranted belief. Sums up Wright:

This line of reply concedes that the best sceptical arguments have something to teach us—that the limits of justification they bring out are genuine and essential—but then replies that, just for that reason, cognitive achievement must be reckoned to take place within such limits. The attempt to surpass them would result not in an increase in rigour or solidity but merely in cognitive paralysis.<sup>17</sup>

Wright proceeds to turn these elusive thoughts into a firm proposal about entitlement. P is a presupposition of a cognitive project if to doubt P would rationally commit one to doubting the significance or competence of the project. Wright then proposes that we have an entitlement of cognitive project when a presupposition, P, of a cognitive project when we have no sufficient reason to believe P to be untrue, and the attempt to justify P would involve further presuppositions of no more secure a priori standing.<sup>18</sup>

Provided then that forming moral judgments is a form of cognitive project - and it is hard to imagine why it wouldn't be - it seems clear that the entitlement of cognitive project also extends to cover moral judgments. The proposition, "My moral intuitions are reliable" is surely a presupposition of the cognitive project of everyday moral judgment,

---

<sup>17</sup> Wright (2004), p 191

<sup>18</sup> Wright (2004), pp 191-2

by Wright's definition of presupposition. Thus, we are entitled to our loose moral judgments.

Let me now turn to contemporary entitlement theories. One of the most straightforward entitlement theories is James' Pryor's. Pryor describes his theory as a form of "dogmatist epistemology," and his description is one of the clearest statements of the commitments of entitlement theory.

The dogmatist about perceptual justification says that when it perceptually seems to you as if  $p$  is the case, you have a kind of justification for believing  $p$  that does not presuppose or rest on your justification for anything else, which could be cited in an argument (even an ampliative argument) for  $p$ . To have this justification for believing  $p$ , you need only have an experience that represents  $p$  as being the case. No further awareness or reflection or background beliefs are required. Of course, other beliefs you have might defeat or undermine this justification. But no other beliefs are required for it to be in place.

Note that the dogmatist is not saying that your justification for believing  $p$  rests on your *awareness* of your experiences. His view is that you have justification for believing  $p$  simply in virtue of *having* an experience as of  $p$ . On his view, your experiences give you justification for believing  $p$ , but it would be misleading to call these experiences your "evidence" for believing  $p$ . For saying that your experiences are your "Evidence" for a perceptual belief suggests that your justification for that perceptual belief depends in part on *premises* about your experiences - as if you were introspectively aware of your experiences, and your perceptual belief were based in some way on that awareness. The dogmatist denies that you need any "evidence" of that sort...

Of course, you can *become* aware of your experiences, by introspection. And your introspective awareness that you have experiences of certain sorts might, together with appropriate background beliefs, provide you with *additional* reason to believe  $p$ . The dogmatist does not deny that. He allows that you may have some justification for believing  $p$  that does rest on your introspective awareness of your experiences, and on background beliefs. He only claims that there is a kind of justification you have which *does not* rest on these things.<sup>19</sup>

In Pryor's account, the dogmatist does not take perceptual beliefs to be *self-justifying*; the dogmatist takes us to be entitled to believe along with our perceptual experiences *without justification*. The dogmatic position is essentially non-reflective.

Pryor claims that only a dogmatist about perceptual justification can successfully escape from the skeptic's challenge. Does Pryor's theory plausibly extend to moral

---

<sup>19</sup> Pryor (2000), p 519



judgments? First, a full theory of entitlements will go beyond perceptual entitlements, and that Pryor's story here is, by his design, incomplete. But even Pryor's account of the simplest of entitlements - our entitlements to perceptual judgments - covers certain forms of moral judgments.

Pryor fleshes out his notion of perceptual justification in the following way: we are justified in trusting "**perceptually basic** propositions, or propositions that our experiences **basically represent** [Pryor's bolding]."<sup>20</sup> Pryor contrasts perceptually basic propositions to those that require background knowledge. The judgment that there is a blue object in front of me is perceptually basic, says Pryor; the judgment that there is a policeman in front of me is not. This is not to say that perceptually basic propositions are merely statements of sense-data, à la some logical positivists. Perceptually basic propositions are "about manifest observable properties of objects in the world."<sup>21</sup> Which propositions are actually basic is actually an empirical matter. Says Pryor, it is up to the cognitive psychologists whether "there is a complete hand" or "there is a facing flesh-colored surface of such-and-such a shape" is the basic supposition. Pryor also admits that what's perceptually basic will vary between people, and will vary for one person over time. I take it, for instance, that, under some descriptions of autism, what's perceptually basic for the one person is the judgment, "the person is raising the corners of their lips" and what's perceptually basic for another person is the judgment, "the person is smiling," or even, "the person is happy." Thus, says Pryor:

---

<sup>20</sup> Pryor (2000), p 539

<sup>21</sup> Pryor (2000), p 539

The official version of my view is that we have immediate *prima facie* justification for believing those propositions that our experiences basically represent to us - whichever propositions those turn out to be.<sup>22</sup>

Pryor's intention here is not to provide a complete account of all possible entitlements, but merely to provide an account of the simplest entitlement: our entitlement to bare perceptions. Surely there will be more entitlements to cover more complex, synthetic cognition. But even here, we can see that even the simplest entitlement extends to the moral sphere. Do our experiences basically represent moral judgments to us? The answer will depend, as Pryor suggests, on the details of each person's phenomenal experience. But it seems that, for many of us, some moral judgments *are* basically represented in experience. These are the sorts of judgments that are sometimes referred to as "raw intuitions." If I see a state of affairs and it strikes me as obviously wrong, then I take it that the judgment is not the result of the application of background knowledge, but, rather, is a part of the basic representation of my experiences. I experience certain events as wrong, or horrifying, or repugnant, or dignified.

Other entitlement theories interface even more easily with moral judgment. A variation on Pryor's version of entitlement theory is Lawrence Sklar's methodological conservatism. Sklar thinks that a fairly weak principle can get us out of various skeptical dilemmas.

What is the principle of methodological conservatism. Basically, the idea is that the very fact that that a proposition is believed can serve as a warrant for some attitude to be rationally maintained in regard to believing it.<sup>23 24</sup>

---

<sup>22</sup> Pryor (2000), p 539

<sup>23</sup> Sklar (1975), p 375

<sup>24</sup> This is Sklar's earliest, and most intuitive, formulation of methodological conservatism. Sklar eventually admits later in the same paper that, so formulated, the principle is too strong. It is very hard to show why

Where Pryor's perceptual dogmatism focuses on perceptual experiences as entitling beliefs, Sklar's conservatism works directly on the beliefs themselves. Sklar's conservatism is even easier to apply to moral judgments, because we don't even have to worry about the question of whether or not moral experiences count as a form of perceptual judgment. Sklar works, so to speak, on the discovered phenomenology of belief, rather than the phenomenology of perception. If I already believe that eating meat is right, and I discover the alternative hypothesis that eating meat is wrong, and there is no greater reason to believe the alternative hypothesis, then I may continue to believe that eating meat is right.

Burge makes the extension of entitlements to non-perceptual faculties explicitly. Burge grants that perceptual beliefs are justified in virtue of the individual's having certain sense experiences, seemingly in parallel with Pryor.<sup>25</sup> But Burge adds that in any form of reasoning that involves multiple steps, we must deploy and trust our memory. Memory plays a "preservative" role. Once we demonstrate to ourselves a particular judgment satisfactorily, memory preserves our acceptance of that judgment for use in future reasoning. But if we reserved our trust in our own memory until we had a satisfying argument, we'd run straight into a vicious circle. We could never produce such an argument, for the use of memory is requisite for any argument. A prior demand for

---

believing *p* would serve as a *reason* to believe *p*. The humbler proposal he eventually commits to is that, *if you already believe some proposition, it is unreasonable to cease to believe the proposition merely because of the existence of alternative hypothesis whose positive warrant is no greater than that of the proposition already believed*. However, the earlier formulation is more comprehensible, and the differences between the formulations is negligible for my purposes.

<sup>25</sup> Burge (1993), p 460

justification, applied to basic faculties of reasoning, will lead directly and inescapably to skepticism. Normal human cognitive life, then, requires more than just perceptual entitlements. Entitlements must therefore capture non-perceptual faculties as well. Burge suggests the following approximation, which he calls the Acceptance Principle:

*A person is entitled to accept as true something that is presented as true and that is intelligible to him, unless there are stronger reasons not to do so.*<sup>26</sup>

Once again, the key entitlement principle is crucially non-reflective. We don't need to accept, understand, or even know of the Acceptance Principle in order to be entitled; one doesn't have to be an epistemologist to be entitled to self-trust. The Acceptance Principle, says Burge, allows us to *accept information instinctively*.<sup>27</sup> The Acceptance Principle, says Burge, also shows that we are a priori entitled to accept the testimony of others for the very same reasons that we are entitled to accept the deliverances of our memory: both are intelligible and present themselves as true. He argues that the fact that a faculty's deliverances seem true and are intelligible are defeasible a priori markers of that faculty's being a rational source.<sup>28</sup>

Burge's extension of entitlement to testimony is part of his overall commitment to a very generous epistemology. Unlike earlier thinkers on the subject, Burge doesn't think we are entitled only to those faculties that are *necessary* for all reasoning. There are many internal faculties that make claims about the world that are not necessary to reasoning in general - our sense of smell, for example, or our ability to read facial expressions.

---

<sup>26</sup> Burge (1993), p 467

<sup>27</sup> Burge (1993), p 467

<sup>28</sup> Burge (1993), p 469

Perceptual judgments are not indispensable to all reasoning, but any reasonable entitlement theory must cover them.<sup>29</sup> "Such resources may enrich reason without being necessary to every rational activity," says Burge.<sup>30</sup>

What may we trust then? Burge suggests that we are entitled to trust rational resources. A rational resource is a resource that's aimed at truth, though not necessarily always correct; but the fact that a rational resource is aimed at truth is sufficient to *treat it, prima facie*, as source of truth. Says Burge, the very fact that a rational resource presents *content* marks it as having an "a priori prima facie connection to truth." This is because "content is constitutively dependent, in the first instance, on patterned connections to a subject matter, connections that insure in normal circumstances of true thought presentation. So presentations' having content must have an origin in getting things right."<sup>31</sup> The very fact that a rational resource presents content requires that it has some patterned connection to the world, and that is sufficient grounds to begin by supposing, a priori, that the resource presents truths about the world.

How then, are we to tell the rational resources from the irrational noisemakers? Once again, we have no sure path to telling a rational resource from an irrational one, suggests Burge, but we have an a priori entitlement to presume that certain claimants are, prima facie, rational resources. Intelligibility is an a priori prima facie sign of rationality, because the presentation of propositional content presupposes at least a derivative connection to a system of perceptual, cognitive, and practical interactions with a world.

---

<sup>29</sup> Burge (1993), p 466

<sup>30</sup> Burge (1993), p 470

<sup>31</sup> Burge (1993), p 471

In short, prima facie intelligibility is a prima facie sign that a source is rational, and a source's being rational is a prima facie reason to take the deliverances of that source as true. Says Burge, intelligibility, rationality, and truth are all of the same domain - cognitive, truth-seeking activity - so prima facie intelligibility confers a prima facie entitlement to presume rationality and so presume reliability. None of these connections are vetted with the force of proof, but, of course, this is a theory of entitlements, not of firm foundations.

This argument is not intended to refute the skeptic, says Burge. Burge is comfortable assuming that knowledge about the external world is possible; what he's looking for is a theory that explains that possibility, and sets out its boundaries.<sup>32</sup> Once we see that something like the Acceptance Principle is the most plausible source of knowledge, we will see its generality, says Burge.

Burge's Acceptance Principle extends easily to cover moral judgments. Insofar as I have the experience as of something's being wrong, or the experience as of some moral axiom's being true - if something seems wrong or true to me - then I am entitled to take the corresponding judgment to be prima facie true. Furthermore, the Acceptance Principle easily extends to moral testimony. Part of the main argument of "Contention Preservation" is that one is entitled to trust testimony on the very same grounds as one is entitled to trust one's own faculties. Testimony is comprehensible and so prima facie it comes from a rational source. Thus I have a defeasible entitlement to presume any comprehensible person rational and so to presume their testimony true. This entitled trust

---

<sup>32</sup> Burge (1993), p 470

for others obviously extends to cover moral testimony, as well. An interlocutor's claims about something's being good or bad are just as intelligible as my own moral intuitions, and thus bear the same sort of prima facie entitlement.

### **Why we should expect any reasonable entitlement theory to cover moral judgment**

I've shown that, for a representative sampling of entitlement theories, each theory extends to cover moral judgment. I've tried to show that these extensions aren't the result of desperate semantic juggling, but are clean, plausible, and even obvious extensions of the entitlement theories. However, I do not take any entitlement theory yet proposed to be entirely successful. I agree with Wright: though no currently available entitlement theory is has entirely overcome its theoretical obstacles, some version of entitlement theory must be correct.

In writing about the extension of entitlement theory to moral judgment, I am, I admit, jumping the gun a bit. I had hoped to provide an argument that *any* reasonable entitlement theory *must* extend to moral judgment, but I find myself unable to do so. The properties common to any possible entitlement theory as so poorly developed and so minimal that such an argument is likely impossible. It is possible that some future candidate entitlement theory will arise that does not extend to moral judgments. But I doubt it, and the fact that all of the present theories so easily extend to cover moral judgments I take as strong evidence in favor of this approach.

Let me close this section with some further thoughts about why it seems likely that any reasonable entitlement theory will extend to moral judgment. The general mood of entitlement theory is one of generosity, of starting in a mood of acceptance rather than one of suspicion. Why ought we think that generosity extends to the moral domain? Many of the entitlement theories we looked at are hyper-permissive and extend to cover loose moral judgments quite easily. Any entitlement theory formulated along the lines of Pryor's or Sklar's or Burge's - that entitles us to accept any proposition that *prima facie* seems right to us, or that we already believe to be true, or that is comprehensible - extends easily. The majority of entitlement theories out there seem to extend entitlements to any propositional claim that carries it with it some minimal stamp of cognitivism - either a claim to truth, or seeming to be right, or prior commitment, or even mere intelligibility. For most of us, our naïve moral experience comes with those stamps of purported cognitivism.

But why should we expect *any* reasonable entitlement theory to have such minimal conditions for entitlement? To understand this, we have to turn to the underlying motivations for turning from more traditional theories to entitlement theories. Entitlement theories are constructed to capture the way in which everyday knowledge works. They begin with a commitment to the reasonability of everyday knowledge. Thus, entitlement theories are likely to have quite minimal conditions for granting entitlements, because entitlement theories must grant entitlements to non-reflective agents - the holders of most everyday knowledge. Entitlement theories cannot, for example, demand that an agent understand entitlement theory, for this would imply that almost all non-philosophers (and



many philosophers) were not entitled to their every-day beliefs. Entitlement theory ought not demand extensive knowledge of the nature, structure, and inner workings of the one's cognitive abilities as prerequisites for being entitled to trust those abilities. This would leave most people at cognitive sea.

Furthermore, it seems that there ought to be some first-personally available criterion for entitlement. To the degree that an epistemic theory like entitlement is a normative theory, there must be some sort of phenomenally available stamp for the legitimate starting points designated by that epistemic theory.<sup>33</sup> Even if it is, as Burge says, instinctive, an agent ought to be able to recognize their legitimate entitlements, even if they cannot fully describe them as such. Furthermore, this phenomenal stamp must be easily accessible and comprehensible, in order to capture the successful knowledge in philosophically unreflective agents. The stamp cannot depend, for example, on some fine distinction between phenomenal states that requires meditative training, or a working knowledge of Husserl. It seems likely that the proper phenomenal stamp will be a very general one.

Furthermore, it is unlikely that a fully developed entitlement theory will limit entitlements to merely those bearing the stamp of perceptual cognitivity - that is, to limit entitlements to propositional claims directly arising from perceptual states. Though many of those working in entitlement theory take perceptual entitlements to be the clearest and best case for entitlements, none think that perceptual entitlements are the only entitlements. In order for everyday cognition to work, we need to trust not only our

---

<sup>33</sup> Dretske a similar point in (2000), pp 595-6.

perceptual abilities, but non-perceptual cognitive abilities, like short-term memory and basic logical abilities.

The possibility remains that entitlements will be limited in some particular way. Perhaps, we might imagine, entitlements will only extend to the cognitive minimum - those faculties necessary for logical and scientific reasoning, which include sensory perception and short-term memory, but excludes the rest. But this, too, would fail to capture so many everyday cases of knowledge. Take, for instance, our ability to judge the emotions of another person; or a therapist's ability to feel out the best thing to say to a bereaved patient; or a military officer's ability to quickly assess, in a moment, the best tactic for a newly evolving situation.

It seems then, *prima facie*, that entitlements ought to be broad-based. It is unlikely that entitlements would extend strictly to very generic cognitive faculties, like general faculties of logic, because there are many very domain-specific faculties, like the empirical senses, memory, and reading emotions, which it seems like entitlement theory ought to capture. It is also unlikely that entitlements will be limited only to concrete, perceptual faculties, since we clearly need entitlements to things such as short-term memory, synthetic judgment, and the like. Thus it seems likely that any reasonable future entitlement theory would also extend to moral judgments.

## **Part II: From Entitlement to Humility**

I've argued thus far that entitlement theory is the most plausible epistemic grounding for loose moral judgments. Entitlement theory shows us that we are reasonable in trusting our loose moral judgments, but we have to trust defeasibly. Our entitled trust is tentative and revocable in the face of future evidence. I argue that, as a consequence of the epistemic analysis, we cannot treat ourselves as morally self-sufficient.

My argument centers around one basic idea: that we are each but one rational agent among many. Each of us has a priori no better or worse claim to epistemic access to the moral truths than other rational agents. Entitlement theory permits us to trust our unsecured judgments and cognitive abilities. Entitlement theory carries with it no grounds for thinking that my access is *better* than that of another rational being - thus, our testimony and theirs ought to downgrade and defeat each other in similar ways. In fact, as a consequence of entitlement theory, the present state of evidence suggests that we ought to think there are other people with reliable access to moral facts, an access comparable to our own. The evidence indicates that we have epistemic peers in the moral domain. This epistemic peerhood is the reason why we cannot be morally self-sufficient

### **Some simplistic arguments for moral self-sufficiency**

I intend to argue this claim against the intelligent supporter of moral self-sufficiency. But first, I think it will be useful to discuss some simplistic arguments for moral self-sufficiency. The diagnosis of these simplistic arguments will set the stage for

the discussion of the sophisticated version of moral self-sufficiency. The sophisticated versions, I will argue, fall prey to the same problems, albeit in subtler formulations.

Here is the first simplistic argument: that our entitlement to trust ourselves grants us the entitlement to treat ourselves as infallible. Suppose that entitlement theory grants that I trust my intuitions. Suppose I have an intuition that  $p$ , and that I have a further intuition that the former intuition is infallible. If I am entitled to trust my intuitions, then I am entitled to trust this meta-intuition's guarantee of the first-order intuition. Thus, I have a reason to take myself to be infallible. Thus I am entitled to believe  $p$ , and to believe that I am infallible in this belief.

There seems to be something very funny about this form of reasoning. It seems like I've bootstrapped myself out of defeasibility. Entitlements can't possibly work this way. There may indeed be a way for agents to rid themselves of the tentativeness and defeasibility that comes with entitlements, but surely it can't be as simple as just adding a second-order intuition. After all, that second intuition is itself merely another intuition. The flaw of this simplistic argument to infallibility is that it treats the meta-intuition itself as infallible. But, since the meta-intuition is itself an intuition held on a mere entitlement, it is also defeasible. Thus, a consideration that bears against  $p$  also bears against my infallibility in judging  $p$ . After all, the second intuition is merely a *claim* of infallibility. Suppose, for example, that there is a new oracle - the Oracle of Seattle. The Oracle of Seattle claims that there will be an earthquake next Saturday. Furthermore, the Mayor of Seattle claims that the Oracle of Seattle is infallible. When the prediction turns out to be wrong, that is evidence both against the Oracle's infallibility and against the Mayor's .

The simplistic argument to infallibility is simply a more convoluted version of the following argument:

Tom: I see a red ball.

Jerry: No, the ball is blue.

Tom: It can't be blue. It's red.

Jerry: Why do you say that?

Tom: Because I see that it's *definitely* red! So it's *got* to be red! So you've *got* to be wrong!

Tom's mistake here is thinking that adding an emphasis creates an epistemic guarantee. The emphasis may arise from a genuinely distinctive phenomenal experience, but the emphasis itself does not provide an iron-clad guarantee. If I have some cognitive ability, A, and then support that with the pronouncement of another cognitive ability, B, then this support is only as strong as the reliability of ability B. As long as ability B is itself defeasible, then we have still failed to make our claims waterproof. Good evidence against the reliability of cognitive ability A will transfer and also function as evidence against the reliability of any cognitive abilities that attest to ability A's reliability. This is true no matter how many further guaranteeing systems we add. If a disagreement weighs against my intuition, it will weigh against any other defeasible second-order, supporting intuitions. So long as the higher-order intuitions are defeasible, I cannot get out of the epistemic effect of disagreement merely by adding higher-order intuitions to the pile.

Let us turn to the second simplistic path to self-sufficiency. Some have argued that it is rationally incoherent to doubt one's own rational faculties, for such doubt is self-defeating. Wright sketches a form of entitlement he calls the entitlement of rational deliberation. He writes: "The generic thought is that since rational agency is nothing we can opt out of, we are entitled to place trust in whatever (we have no evidence against and which) needs to be true if rational decision-making is to be feasible and effective."<sup>34</sup> This form of entitlement yields an interesting approach to self-sufficiency. I cannot opt out of rational deliberation; thus, I cannot doubt any faculties necessary for my basic functioning. It's impossible to reason my way into doubting my basic faculties of reasoning, for such an argument relies upon the reliability of those very rational faculties in order to be convincing. Thus, the core rational faculties are immune to defeat, on pain of incoherence. Core rational faculties get a special bye from the requirement of defeasibility.

This argument from internal coherence has more legs than the previous argument, but I do not think it will carry us all the way to moral self-sufficiency. First, internal coherence considerations only extend to those faculties necessary for coherent reasoning. They will not extend, for example, to perceptions of the external world. It is entirely possible to be a brain in a vat and be a rational decision-maker. Surely, moral cognition is not necessary for the minimum of coherent reasoning. Both perceptions of the outer world, and moral intuitions, are products of faculties that claim to be about *external*,

---

<sup>34</sup> Wright (2004), p 128

*mind-independent facts*. Thus, moral judgments are excluded from these the protections afforded by internal coherence considerations.

More importantly, internal coherence cannot block all suspicion, even towards faculties associated with basic reasoning. The argument from internal coherence must be something along these lines:

1. I have some degree of belief in my rational faculties,  $f$ .
2. I have some degree of belief that my rational faculties are unreliable,  $g$ .
3. Since my argument against my faculties depends on the use of my faculties, it cannot be that the confidence in  $g$  exceeds the confidence in  $f$ . It must be that  $f$  is greater or equal to  $g$ .

But this proves only that I can never believe less than 50% in my rational faculties, based on my rational faculties. It does show, for example, that I can't downgrade my belief in my rational faculties to 0%, based on an argument using my rational faculties. My confidence in the conclusion comes from my confidence in its source, and believing on the basis of my reasoning faculties that my reasoning faculties were more untrustworthy than not would lead to a version of the liar's paradox. But I can certainly downgrade my rational faculties, so long as I maintain  $f > g$ . Though I can never give up on my rational faculties entirely, I can certainly come to doubt them to some degree. This version of the worry from internal incoherence provides a very interesting result: it shows us an upper

bound to the self-doubt we can have in our rational faculties. But this worry cannot forbid us from self-doubt inside that upper bound.

Crispin Wright argues that this entitlement gets going because it's a presupposition for an activity that we can't opt out of – the activity of rational deliberation. But I can surely opt out of the activity of *rational deliberation with absolute certainty in my faculties*, and opt into the activity of *rational deliberation with some degree of doubt towards my faculties*. Considerations of internal coherence can only grant immunity from downgrading for activities I can't opt out of, but surely I can opt out of cognizing with absolute certainty.

Thus, neither the argument from infallibility, nor the argument from internal coherence, will get us to moral self-sufficiency. But their problems will recur in more subtle form.

### **The argument from trumping**

Let me now present what I take to be the most plausible and significant argument for moral self-sufficiency. It is very hard to imagine life without any moral trust whatsoever - there would be no moral education, no moral advice; the most plausible versions of moral self-sufficiency must therefore make allowances for some circumscribed uses of moral testimony.

One promising solution for moral self-sufficiency would be to grant that an agent may use moral testimony when the agent has no considered judgment of their own on the



topic, but claim that once an agent develops their own considered judgment, their judgment blots out the weight of moral testimony. After all, all that's required to capture many of our intuitions about legitimate uses of moral testimony is to acknowledge that moral testimony has *some* weight, but this does not have to be very much weight at all. The proponent of moral self-sufficiency might plausibly hold that moral testimony is just barely weighty enough to move me to action when I have no substantive moral judgments of my own, but too trivial to have any effect once I form a moral judgment of my own. There are many such epistemic featherweights in intellectual life: a vague memory, or the hint of a feeling, or testimony from an unvetted source. Imagine, for instance, that I have no idea which direction to go to get out of a maze. I see some graffiti saying, "This way!" Now, I have very little reason to trust this graffiti, but given that I have no sense of my own which way to turn, the whisper-thin weight of evidence provided by the existence of the graffiti provides a reason to go in a particular direction, and if I have absolutely no other reason to choose one way over the other, then that whisper-thin reason might be decisive. But if I have *any* more solid reason - the memory of a map, the feeling of fresh air from a certain direction - this counts far more than the graffiti of unknown origin. I have a reason to act according to the featherweight reason only so far as I'm lacking anything better, but any substantive evidence will effectively wipe out the epistemic effect of the featherweight.

Let's call this the *argument from trumping*: that, though moral testimony can have some evidential weight for an agent, the presence of an agent's own moral judgment is of sufficiently greater importance or weight as to *trump* any weight from moral testimony.

This is similar to how one might use movie reviews or restaurant recommendations. If I know nothing about a restaurant, then the fact that somebody recommended the restaurant is a reason to think the restaurant good. It is entirely reasonable to act on the testimony - for example, to actually go to the restaurant. But, it seems, once I go to the restaurant and eat some meals for myself, then my direct judgment supercedes any testimonial reasons. The fact that others like it cannot weigh on my judgment in any way - in fact, what I'm most likely to do is lose trust in anybody that recommended such a lousy place.<sup>35</sup>

This argument from trumping is particularly compelling in the space of moral judgments because of the special relationship we seem to have with our moral judgments. We have no special relationship to our empirical judgments; there is nothing personal about eyesight. I see reasonably well, but I will happily admit that some others see better and others see worse. More importantly, I have no trouble trusting another person's vision over mine, if I have evidence that they are more sharp-eyed or more color-discerning or the like. Moral judgment seems somehow more personal. Even though we are often willing to describe different agents as more or less apt at moral reasoning, it seems vital that our own actions and beliefs emanate from our own moral judgments.

The argument from trumping is a very plausible defense of moral self-sufficiency. It explains away most normal uses of moral testimony while still dismissing moral

---

<sup>35</sup> In fact, I think this view of aesthetic recommendations and aesthetic self-sufficiency, though common, is flawed, for exactly the same reasons that moral self-sufficiency is flawed. I bring up this example only because this attitude is familiar, not because I think it is correct.

disagreement. In order to resist the argument from trumping, I will have to show that moral testimony provides *sufficiently weighty reason to effect my own beliefs*.

### **The existence of moral peers**

To block the argument from trumping, I will have to show that contrary moral testimony is of the *right kind* and of *sufficient strength* to have some non-trivial effect on my belief-state. I will argue that we have, in most circumstances, reasons to accept certain others as our moral peers, and so to accept their testimony as being of sufficient strength. I define a "peer in domain X" as an agent whose believing p gives me a *substantive reason* to believe p. Defined this way, the argument against self-sufficiency breaks down into two steps: first, showing that there are moral peers, and second, showing that a substantive reason to believe -p, when I have judged p, ought reduce my confidence in p. Let me differentiate the notion of a "moral peer" from the broader notion of a "moral source." A moral source is somebody whose belief that p gives me some reason to believe p; that reason could be substantive or featherweight. Moral peers are particularly important kind of moral source. The argument from trumping gains its strength by allowing that there are moral sources, while denying that there are moral peers.

Let us begin by considering the existence of moral sources. It does seem, at first glance, that there are moral sources. The examples are hard to find when we move to abstract principles, but examples come readily to mind when we look to more concrete

and mundane affairs. We ask for assessments of other people's moral character; we trust our friends' judgment about who is compassionate and who is cruel. Annette Baier argues that normal human life would be impossible without some moral trust. She points out, for instance, that social cooperation requires that I delegate judgments to others.<sup>36</sup> I might wish contribute money to a good cause and pressed for time, ask a friend which cause I should contribute to. I am in this sense taking moral testimony - I trust not only their assessment of the morally neutral facts, but also their moral assessment of those facts. When I use a newspaper's voting guide to guide my hand for the vast array of city ordinances, I am trusting the author's research capacity and her moral judgment.

All of these examples, however, are cases where I use testimony where I have no judgment of my own; they support only the existence of moral sources, and not moral peers. The mere existence of moral sources is compatible with the featherweight view of moral testimony, and compatible with the sophisticated moral self-sufficiency I've described above. But there are plenty of everyday examples that suggest the existence of moral peers. Take, for instance, any situation in which a friend urges me to reassess a situation because she suspects that I am biased, or blinded by rage, or unable to assess the situation neutrally. I might, for instance, after having been insulted by a colleague, angrily rush to report it to our boss. Suppose, though, that my friend tells me to wait. I'm too angry to consider the issue fairly right now, she says, and need to cool off. And, she says, I have always disliked my colleague for morally irrelevant reasons - they have a grating voice - and I am letting that color my moral judgment. When I take my friend's

---

<sup>36</sup> Baier (1986), p 232

advice to heart and stay my hand, I am treating her as a moral peer. Even if I don't see for myself in the moment that she is right, even if my own argument still seems to me to be sound, even if I don't actually think that I am biased at the moment, the possibility raised by her testimony is sufficient to alter my degree of confidence and my course of action.

These examples are not intended to be conclusive; I am aware that determined opposition could suggest alternate explanations for these cases. The examples are only intended show the existence of moral peers, rather than being an alien thought, fits with many familiar social phenomena.

It is, however, interesting to note that in these everyday examples peer testimony reduces, but does not obliterate, the confidence I have in my own moral judgment. Most of the really intuitively plausible cases of moral peerhood seem to share this quality. Our intuitions comport with the existence of moral peers, but also seem to set an upper bound on the impact of their testimony. The presence of such an upper bound is compatible with my own conclusions; I aim only to show that moral disagreement has some impact, not that it has unlimited impact, nor even to show that moral disagreement has as much weight as my own moral introspection. I suspect that there are very good reasons for such an upper bound. Cases of treating another person as a moral expert - that is, of treating them as so vastly superior to myself that their moral judgment trumps my own - sit uneasily with many intuitions. I suspect that we will discover very strong considerations against treating others as moral experts. I am happy, then, to grant that there is, in this respect, some disanalogy between moral judgments and empirical judgments. There may be something wrong with trusting another's judgments *to the complete exclusion of one's*

*own judgments* in the moral sphere, and perhaps that there is something wrong with trusting another's judgments as much as one's own. But there is also something wrong with excluding testimony and disagreement entirely.

### **The generalization argument**

I will now argue that, given that we are merely entitled to our moral self-trust, we ought to trust some other agents. This argument is the lynchpin of my positive project. In brief: we ought to regard as peers those who have, as far as we can tell, approximately the same cognitive capacities as we do. This is because the mechanisms which underlie the reliability of my own moral judgments are, to the best of my knowledge, shared by other human beings. Thus, the entitlement I have to trust my own mental abilities extends to trusting the mental abilities of others.

This argument applies to all of our entitlements, moral and non-moral alike. If I am to trust my own vision, I ought to substantively trust the vision of (some) other people, for their visual equipment is, to the best of my knowledge, similar to mine. For the same reason, if I am to trust my own moral judgments, then I ought to extend a similar order of trust to the moral judgments of those who have similar cognitive equipment to mine.

My goal here is to show that the generalization argument, which is uncontroversial for empirical matters, must extend to moral cognition. I claim that *any entitlement theory that licenses thinking of myself as prima facie rational requires that I*

*also that I think of others sufficiently like me as prima facie rational.* I mean by "rational" here that the rational agent or ability reliably tracks the truth in the relevant domain. (One does not have to track the truth perfectly to be rational, under this definition.) Purported rationality is a prerequisite for the existence of moral disagreement. In order for an agent to take herself to be in a disagreement, the agent must believe her relevant judgment to be of an objective affair, and be moderately confident in that judgment. The agent must, at the very least, take herself to be making a claim about something distinct from her personal preferences.

Merely from these prerequisites to disagreement - the objectivity of content and self-confidence - we can already make some substantial deductions. I am entitled to believe that my loose moral judgment is likely true. Since the content of that judgment is an objective claim, I must take myself to have some ability to reliably track objective, mind-independent facts. In trusting myself, I am imputing to myself some reliable ability for tracking the truth - for getting onto states of affairs outside of me. Insofar as I take my judgment to be *evidential* rather than pragmatic - to be about what is actually the case - then it must be that the abilities and methodologies involved in generating these judgments are reliable. That reliability is the explanation for the purported correctness of my judgments. Entitlement theory says I am reasonable in imputing to myself such a reliable ability, even lacking an account of why the ability is reliable.

Now we turn our gaze to other people. Ought we trust them? Yes, for trusting others follows from trusting ourselves. In the situation of entitled moral judgment, we are trusting that our cognitive capacities will be reliable, but we have little to say about these

cognitive abilities. We don't know what they're like, how they function, the basis for their reliability; we don't even know, in many cases, which cognitive abilities we're using. But we do know that certain other people - our peers - are cognitively very similar. They seem to have mostly the same abilities; we seem to be able to understand each other and reason very similarly in most situations; we seem to have the same physical apparatus, and we seem to come from similar sources - similar genetic material, a similar evolutionary background.<sup>37</sup> Given that we have no particular account of where in our mental makeup our moral truth-tracking ability resides, and many reasons to think that the mental makeup of others is largely like our own, we have *many reasons in favor of taking others to be our moral peers and few reasons against*.

Any entitlement that licenses me in thinking of myself as rational in some domain thus requires that I think of others as rational, provided I have reason to think that they are sufficiently like me. To the extent that I take my judgments to be accurate, I must take myself to have some ability to track the truth. That ability depends on my rational features and abilities. Others who have similar features and abilities are then likely to have similar capacities for truth-tracking. What counts as "sufficiently like me"? It's very hard to say because of the impoverished nature of our epistemic self-knowledge. In the situation of entitlement, I don't have enough information to draw the line precisely but, for my peers, the evidence weighs on the side of similarity. For my peers, there are many empirical reasons to think that they're similar to me, and no prima facie categorical reason to think they're not. There are, of course, sometimes contingent reasons to think that

---

<sup>37</sup> This argument should also work for a monotheist creationist, who, I take it, takes all humans to have been created by the same God.



certain people aren't as truth-tracking as I am - they're biased, they're angry, they haven't thought things through as long as I have, they don't have the background or education I have in this particular area. But the proper initial presumption, given that they're people rather like me, is that they're also rational like me, and thus approximately as morally reliable.

My entitlement to trust my own mental abilities extends to trusting mental abilities of others, because there is *a wide range of empirical evidence that they are creatures substantially like me*. They are human, they seem to have brains similar to mine, educations similar to mine, genetics similar to mine. They seem to reason largely as I do. They're of the same species, and have been shaped by the same evolutionary pressures.<sup>38</sup> Since I take my own faculties to be providing me with substantive reasons to believe and since, for some other agents, I have every reason to believe that their faculties are like mine and little reason to suspect they are unlike, then I ought to believe that their faculties are also well-functioning, and thus should take them as providing substantive reasons to believe.

The argument I've given depends on a thick, juicy stack of empirical data. It's not an argument that the bare existence of testimony or the bare existence of other speakers gives us reason to believe. It depends on some very contingent empirical data: that we're surrounded by people that all of our experience, history, and science tells us are largely like us. I don't think, for example, that a person stranded on Easter Island, surrounded by

---

<sup>38</sup> This is the version for the believer in humans as evolved creatures. I believe, though I'm not sure, that theists would also have a similar belief in essential similarity, though it would come from something like sharing a creator and being made from the same mold, so to speak.

statue heads of unknown origin chanting, "Chicken is evil! Chicken is evil!" has any reason to trust those statues. It's certain *people* that we have a reason to trust.

Thus, I claim, the argument from trumping is not a viable route to moral self-sufficiency. The argument from trumping depends on claiming that the evidential weight of testimony is of a radically lesser strength than the evidential weight of my own reasoning. But the situation of entitlement doesn't support such an epistemic asymmetry. Rather, in the situation of a disagreement over entitled judgments, the evidence weighs on internal reasoning and external testimony weighing in at approximately the same magnitude. Things might be different if we had something like a proof for our conclusions, but at present, we do not. So long as we are relying on the substantive conclusions of defeasible faculties, and so long as the evidence weighs in favor of the faculties of others being substantially like those faculties, testimony counts and thus disagreement counts.

The argument I've given clearly applies to empirical abilities, but why should we take it to apply to moral abilities? I think that, in fact, the argument applies more strongly to moral judgment. The crux of the argument from generalization is that *given that I don't know very much about the source of my reliability, general features of cognitive similarity weigh more*. The argument applies most powerfully to our entitled moral abilities, precisely because they are the most mysterious of our purportedly objective faculties. We don't have to rely on general features for assessing vision because vision is so well understood, and we have a well-developed and trustworthy method for assessing the reliability of visual systems. It is when we lack such an account that we must default

to general cognitive similarity, and our account of moral cognition is, at the moment, terribly impoverished.

As clear and obvious as some of our moral experience may be, we must admit, on further consideration of the nature of our epistemic relationship with moral facts, that we are each but one agent among many, each making claims about essentially public data. None of us has any essentially privileged relationship to moral reasons, though we may be more or less able or accurate. We are each merely one cognitive agent among many, all looking at the same set of facts.

### **The publicity of moral reasons**

What I'm claiming here is that moral reasons are essentially public, and that this publicity is woven throughout the structure of our epistemic commitments in situations of moral disagreement. Why do I take moral reasons to be basically public? It is an essential part of the phenomenal character of most moral intuitions. When we take another to be *responsible* for behaving a certain way, we must think that the reasons are in principle available to them. We couldn't hold another to be responsible for understanding and following considerations that were essentially inaccessible. Insofar as I judge not only that Tim *ought* to give some of his money to the relief effort but also that Tim is *responsible* for giving money, I must think that the relevant demand is intellectually accessible to Tim. This follows straightforwardly from the "ought implies can" principle. I cannot hold that Tim is responsible for knowing that he ought to do such-and-such, and,

at the same time, hold that the fact that he ought to do such-and-such is a fact only accessible to me. In order to hold others morally responsible, I must believe that they potentially have access to moral reasons. Since our intuitions about moral responsibility commit us to the publicity of moral reasons, then we must take up that publicity, with all its consequences. We must admit that others have, in principle, the same sort of moral access we do. Publicity is a two way street.

The epistemic publicity of moral reasons is quite intuitive; for straightforward moral truths, we take each other to be responsible for discovering them. It is wrong to cruelly insult others without cause. I do not take this to be a private fact, but a public one. I do not take myself to be the only agent in the world privileged with this knowledge; rather, I expect others to be familiar with this fact, and when they are not, I am surprised and think that they have failed in their responsibilities of reflection and moral awareness. I may have some personal relationship to my moral judgments, but insofar as I take my moral judgments to be of objective, public facts, then I cannot deny that others also have access.

What about topics on which one is an expert? One can surely take some fact to be public in principle and yet think that one has better access to that fact than one's interlocutor. Suppose, for instance, that I have an art student, who I hold responsible for learning to tell the difference between red ochre and burnt umber. They take a test on color discrimination and fail, and so I can reasonably judge myself to have better access, and so dismiss (at least for the moment) my students' claims about which earth tone is being used in a particular painting. But this dismissal isn't based on an essential,

insurmountable difference in principle between my own judgments and the testimonial judgments. It is a particular dismissal in a particular context, about a single considered epistemic relationship between myself and my student and our (evidentially supported) degrees of competence in this domain. It is the situation of *discovering that somebody is not a peer*. Defenders of moral self-sufficiency cannot avail themselves of such a situation. In order to show that dismissal is reasonable *between peers*, one needs a principled way set my own judgment as categorically distinct from, and above, the judgment of all others. But such a commitment is incompatible with basic moral commitments concerning the responsibility of others, and the publicity this entails. That we may hold others morally responsible entails that we be morally humble.

### **Kant and me**

The moves of the generalization argument are very much in the spirit of a Kantian transcendental deduction. I argue that in order for moral knowledge to be possible, I must be entitled to trust my moral faculties. It follows, I argued, that if I take myself to be reliable, it must be in virtue of my cognitive capacities and abilities. To the extent that other people share those human features, I ought to take them to be reliable also. This argument loosely follows the form of a Kantian deduction to the value of other people. To oversimplify the Kantian deduction: I take my will to be authoritative, therefore I must take my will to be valuable, therefore I take wills to be valuable in virtue of its

rationality, therefore other people with rational wills are valuable.<sup>39</sup> The key step in both is claiming that if I assign myself a certain property, the only reasonable explanation for my having that property is in terms of certain other properties, which are shared by others. But if I'm availing myself of Kantian deductions-type arguments for my rather slender and tenuous argument for some minimal degree of caution and humility, why not go whole hog and avail myself of the full-strength deduction for value, and get a robust, Kantian ethics out of it?

I think that the epistemic deduction is more plausible. My epistemic deduction far has fewer commitments, and is thus more universally palatable, than the value deduction. It's plausible to claim that the reliability of my cognitive faculties depends on features that are shared by other cognizers. On the other hand, there are many reasons to think that my *valuing* my own rational will depends on some essentially personal feature of it: namely, that it's *mine*. The objectivity of value is much easier to deny than the objectivity of basic epistemic norms.

For a subjectivist, the Kantian deduction would be unconvincing. The subjectivist would simply deny the premise that I valued my will for impersonal reasons; they might hold that I valued my will because it was *mine* and that such a personal basis couldn't be extended to others' wills. On the other hand, it's very hard to think that I trust my rational faculties simply because they're *mine*. The reasons that underlie reliability have to have something to say about the way in which the faculty hooks up, gets onto, and processes data from the external world, and does it well. This is because the function of rational

---

<sup>39</sup> This is an oversimplification of the version presented by Korsgaard (1996).

faculties is essentially outwardly directed – they're supposed to get on to some external, objective truths. It's very hard to deny that there are objective norms for proper epistemic procedure.

It is hard to deny that epistemic procedures for getting on to objective facts are, themselves, objective. Many deny the objectivity of value, but few deny the objectivity of good truth-gathering methods, even the subjectivist. It seems like a presupposition of any argument, including a subjectivist argument, that there are objective epistemic norms.

The subjectivist's commitment to subjectivism is itself an objective claim. The subjectivist doesn't think subjectivism is correct because of something special about her relationship to the theory; she thinks it is correct on objective, rational grounds. It is much easier to be a radical subjectivist about value than a radical subjectivist about truth. Since my argument depends only to commitments to objective norms of epistemology, its reach is broader.

### **Problems from the non-reflectivity of entitlements**

I now take myself to have presented a complete, positive argument for trusting others, based merely on the fact that one must take oneself to be entitled to one's moral judgments. But a problem arises from the very structure of entitlement theory itself. Entitlement theory must, by its very nature, grant entitlements to non-reflective agents. Does my argument demand a greater degree of reflectivity than entitlement theory requires?

My argument does in fact depend on a certain degree of reflection in the agent. I deduced, from the fact that we trust our moral judgments, that we must trust our mental abilities used in forming these moral judgments, and those abilities exist in others. In order for my argument to have any effect on the actions of an agent, it must be that the agent recognizes that there is a degree of entitled self-trust they have towards their own judgments. This is reflective awareness. Without it, the agent would not be compelled to accept that others were as reliable as herself, even if that were true.

The rest of my generalization argument proceeds from this claim of reflective self-trust. But, under any plausible entitlement theory, extensive reflective self-knowledge cannot be a pre-requisite for possessing entitlements. Entitlement theories are supposed to fit our intuitions about ordinary knowledge by ordinary agents. In order for ordinary knowledge to be possible, entitlement theories must offer entitlements with some degree of immediacy. If it seems to me that there is a red apple in front of me, then I am entitled to believe that there is a red apple in front of me. I do not need to have any further supplementary beliefs about my visual system, my cognitive processing capacities, nor do I need to be familiar with entitlement theory or any other philosophical theory. An agent cannot be required to know something about, for example, their visual system, or entitlement theory, to be entitled to self-trust. Thus, says entitlement theory, it must be reasonable for agents to believe and act without reflective epistemic self-knowledge.

Many earlier epistemic theories allow for unreflective starting points, but require that they be minimal and very abstract; they were, perhaps, modeled on axiomatic



mathematical knowledge. But in order to capture our intuitions that non-philosophical agents can come to know things about everyday objects, modern entitlement theories typically extend entitlements to a very large body of discrete phenomena - sights, sounds, memories, fuzzy judgments. Thus, entitlement theory endorses not only non-reflectivity in principle, but endorses the simple, everyday unreflectiveness of concrete, particular judgments. Though it is possible that some entitlement theories might be formulated as lending entitlements to cognitive faculties, most leading candidate theories lend entitlement to unreflective judgments of particulars.

In light of this wide-ranging unreflectiveness, we might worry that the generalization argument won't get its hooks into most agents. The generalization argument depends on the premise that an agent is entitled; for an agent to accept the generalization argument, she would have to reflectively acknowledge her own status as entitled. But entitlement licenses unreflective confidence; the agent may be entitled without being rationally required to acknowledge that entitlement. Thus, even if the generalization argument is true, it wouldn't convince reasonably unreflective agents.

This is, I grant, a serious worry. There are two ways out - a tidy way, and an inelegant, brute force way. The tidy way would be to show that the very existence of disagreement itself forces one to reflect on the epistemic basis of one's self-trust. I can demonstrate that this is true under some particular entitlement theories. For example, Burge claims in "Content Preservation" that we are defeasibly entitled to enter as a premise in our reasoning any comprehensible propositional phenomena, including visual phenomena, remembered phenomena, and more cognitively complex phenomena of

judgment. This entitlement, explains Burge, comes from the fact that we have an a priori entitlement to trust the deliverances of any faculty aimed at representing the world, and we have an a priori entitlement to trust that a faculty is so aimed if it gives us comprehensible deliverances. Thus, from the very same principle of entitlement, we are entitled to trust both our own judgments, and to trust testimony, since both are comprehensible phenomena. Since we are entitled to enter them as premises, then in any case of disagreement over a judgment, we can directly enter contradictory premises. Surely, if anything is grounds for defeat, it's a logical contradiction. Thus, under Burgean entitlement, the existence of disagreement defeats my entitlement to self-trust and to trusting others. In order to get either back, I must get past my entitlements and provide some theoretical basis for self-trust; I must begin to reflect. That reflection will surely involve reflection about the reliability of my faculties, which will set the stage for the generalization argument.

I had hoped to provide a generic argument along these lines, but such a generic argument has not been forthcoming. I suspect that it is the case that in any reasonable entitlement theory, the existence of disagreement will trigger the reflection required to get the generalization argument up and running. I also suspect, however, that this argument will only be furnished piecemeal, against the background of specific, well-developed entitlement theories, and that no generic argument along these lines is possible.

In its stead, I will now provide an inelegant, brute force, but properly generic argument. Entitlement theories show that we are not required to justify our immediate

judgments with further reflection. They do not show that we can *ignore* convincing evidence that has been presented to us. A theory that allowed us to ignore contrary evidence that had been raised would not be a theory of *defeasible* entitlements. That is, entitlement theories show that I am not required to reflect on my cognitive resources in order to reasonably believe in my immediate judgments. They do not show that I can ignore manifestly true claims about my cognitive resources when I am presented with them. So, though it does not follow from entitlement theory that an agent is required to begin reflecting about their cognitive resources, it does follow from generic rules of rationality that an agent is required to believe arguments that are manifestly true once these arguments are presented to them. So, I claim, though it may be that not every entitled agent is rationally required to believe the generalization argument, it is rationally required for every agent that has read this chapter.

### **From disagreement to doubt**

I have set up all the pieces; now let me assemble them from across the last two chapters. Here is the complete argument that moral self-sufficiency is wrong, and that, in the face of moral disagreement with peers, we ought to reduce our confidence in our own judgment to some degree; we ought to be morally humble.

In Chapter 1, I showed that there are such things as moral peers; and, as a simple matter of (admittedly contingent) fact, most of us are in a position to discover many of them. Let us assume that we are discussing an agent who has discovered evidence in

favor of there being moral peers, and that she has disagreements with some of those peers. These disagreements will most likely be over specific topics, embedded against a largely shared set of moral values and principles. I do not know of anybody I rate as a moral peer who thinks that torturing children is right, but I do know of plenty of people who I rate my intellectual and moral peers - good, sensitive, thoughtful people - with whom I disagree on issues like vegetarianism, cultural traditionalism, the relative importance of environmental concerns, and the degree of one's special obligation to one's family members.

The most compelling cases against moral self-sufficiency will be disagreements with peers over the sorts of topics over which peers might plausibly disagree. I will focus on those best cases, since what I'm doing here is a proof-of-concept. I am arguing against the position of moral self-sufficiency by asserting that disagreement is the right sort of thing to matter; thus, I confine my discussion to the best cases for moral disagreements' mattering. Imagine, then, a disagreement between two agents who are, to the best of their knowledge, as reasonable, as rational, as reflective, and as morally sensitive as one another. For all imaginably relevant cognitive attributes or experiences, imagine that these agents share them: they have similar intellectual abilities, educational backgrounds, etc. etc. Furthermore, since we often assess one another's intellectual abilities by looking at the degree to which they agree with our confident judgments, imagine that the two agents agree with each other on the vast majority of moral conclusions. Their disagreement is over one particular thorny issue. Furthermore, imagine that the disagreement is longstanding and intractable, for reasons given in Chapter 1.

These specifications may seem to narrow the scope of my arguments, but I do not believe this to be a serious problem. I don't claim that *only* cases like this are important; rather, I claim that these sorts of cases represent the simplest and clearest cases for significant moral disagreement, and are the least complicated to handle. In order to ascertain the impact of less ideal cases, we will have to clarify a number of further issues, including the ways in which we rate people moral peers, the degrees of possible peerhood, and the interaction between the degree of an interlocutor's peerhood and the weight of their testimony. Very little work has been done on the topic, largely because I believe most contemporary moral philosophers have, implicitly or explicitly, subscribed to some version of moral self-sufficiency. By showing through a best-case study that disagreement is the sort of thing that matters moral peer, I hope to motivate further investigation into the theory of moral peers.

Second, though these best-case scenarios are rare, they are precisely the sorts of cases that are intuitively compelling. It is disagreements with respected thinkers, with the very morally sensitive and reflective, with our trusted comrades, which naturally startle us and force us to worry and reflect.

Imagine, then, that one disagrees with a best-case moral peer over a moral judgment, *p*. Let us say that that one judges, after careful consideration, that *p*, and one discovers that one's respected interlocutor judges, after careful consideration, that not-*p*. Let's assume that we have every reason to believe that the interlocutor both genuinely believes not-*p*, and that the interlocutor has actually given the matter due consideration. I presumably have a substantive reason or reasons to believe *p* that underlie my judgment

that p. I now have a substantive reason to believe not-p, namely, that my moral peer so believes. Now the question is, does the substantive reason to believe not-p give me a reason to change or reduce my present degree of belief in p?

It might strike one that the matter is settled by the generalization argument. What I've argued in this chapter is that entitlement theory is the most plausible epistemic theory covering our self-trust in our unsecured moral judgments. I have argued that it follows from this start in entitlement that our self-trust generalizes to trust in others. This means that we ought to treat others as having a substantive degree of reliability. The generalization argument shows that the testimony of a moral peer is substantive *on the same order* as the reasons generated by my own judgments, because, as far as we know, the mechanisms generated the reliable judgments are the same. This, I think, is the nail in the coffin for moral self-sufficiency. It seems, at least *prima facie*, that when I judge p and then encounter a substantive reason to believe not-p, I ought to reduce my confidence in p by some degree.

There is, however, still the standing possibility that a substantive contrary reason to believe not-p will not necessarily constitute a reason to change my degree of confidence in p. But this possibility seems *prima facie* very strange; it seems a general rule of rationality is that a substantive reason bearing on the truth of some proposition, p, ought to change my degree of confidence. I will consider this possibility at greater length in the next chapter; but it strikes me that, in this case, the burden of proof is on defender of moral self-sufficiency to show why an agent could be permitted to ignore a substantial reason to believe. I raised some reasons earlier why one might dismiss testimonial

reasons: that testimonial reasons might be of trivial strength, or of a different order of evidentiality. But these considerations have been defeated by the argument from generality. The argument from generality shows, from the nature of entitlement and the weight of empirical evidence, that other human beings generate reasons from faculties that are prima facie as reliable as our own, and thus of a similar order of evidential strength. I claim that it is prima facie true that substantive contrary reasons will always give one a reason to downgrade one's belief; I will consider particular positions against this prima facie principle in the next two chapters.

I take the positive part of my argument to be finished; the discussion of peers, entitlement theory, and generalization show completes the basic thought as to why we ought to treat some others as sufficiently trustworthy such as to allow disagreement to have some effect on us. There are several remaining objections to deal with, significantly more sophisticated than the arguments from trumping I discussed above. The sophisticated arguments will be cognizant of the usefulness of disagreement in many empirical reasoning. Thus, the plausible objections will insist on some form of disanalogy between empirical disagreement and moral disagreement. I will spend the remainder of this dissertation discussing those sophisticated objections.

## Chapter 3

### The Epistemic Case for Dismissal

Given the current range of evidence, moral disagreements can seem more threatening than empirical disagreements. Though we are entitled to both realms, our empirical abilities are much more satisfactorily corroborated. There is widespread empirical agreement on most everyday judgments; furthermore, using our empirical abilities we've managed to find a very convincing theory that explains the reliability of our empirical abilities. We have far less corroboration in the moral realm, and we have no satisfying theory explaining our purported moral abilities. The widespread existence of moral disagreement ought to have some significant effect on our moral self-confidence. But many people's intuitions, and the position of moral self-sufficiency, reject this: they claim that moral disagreement is epistemically unimportant.

In the previous chapter, I've argued against moral self-sufficiency by arguing that, in any situation of entitlement, we must regard the judgments of our peers as providing significant evidence. This is because we are each only one fallible cognitive agent among many similar agents. But merely showing that moral testimony is significantly evidential is not sufficient to show that disagreement ought have some effect. I must also hold that, in the situation of moral disagreement, *significant evidence that not-p gives one reason to downgrade one own belief that p*. Let me call this the *open downgrade principle* - that my belief that p is open to influence from significant evidence to the contrary. I claimed in the end of the last chapter that the open downgrade principle was intuitive, and that the



burden of proof was on my opponent to refute it. There are several notable opponent positions that refuse this step in the case of moral disagreement. These opponents grant that moral testimony provides significant evidence, but refuse to grant that this significant evidence ought influence our moral beliefs. They are denying that downgrading is entirely open; they seek, at the very least, an exception in the case of moral testimony. In this way, an opponent could grant my arguments from Chapter 2, and still hang on to moral self-sufficiency.

I will present only a very brief argument for the general form of the open downgrade principle - it seems obvious, and no writer seeks to reject the principle entirely. Given that the generalization argument from the previous chapter establishes that contrary testimony gives me substantive evidence, and assuming that the open downgrade principle is true, then we have vanquished moral self-sufficiency. The defender of moral self-sufficiency, then, in order to hold her ground, must insist that open downgrading does not apply to moral cases. They must seek an exception from the open downgrading principle for moral disagreement. I take this disanalogy to be the most plausible remaining path to moral self-sufficiency, and will spend the remainder of this dissertation discussing various forms of this response.

In this chapter, I will examine epistemic arguments for the disanalogy between moral and non-moral disagreement; in the next chapter, I will look at arguments from moral grounds. I will devote most of this chapter to a consideration of Thomas Kelly's epistemic arguments for self-sufficiency.

Kelly argues that, in situations of absolute epistemic parity between two interlocutors, contrary testimony from another a peer does not transmit. Instead, contrary testimony only gives an agent a reason to stop rating their disagreeing interlocutor a peer. Kelly's argument applies to moral cases especially, as Kelly himself notes, because, though the situation of absolute epistemic parity is quite rare in empirical cases, it is the paradigmatic case of moral disagreement.

I will argue, contra Kelly, that the evidential nature of moral testimony carries through to situations of disagreement. I will argue that we may not dismiss the testimony of others in the moral case *precisely because our moral abilities are defeasible* - that is, precisely because we have them on an entitlement. I will claim that dismissing the opponent depends on a presumption of infallibility in our own cognitive abilities - an unreasonable presumption under entitlement.

## **Part I. How vulnerable are our entitlements?**

Before we begin a consideration of Kelly's arguments, I'd like to further consider some of the concepts in play. First, the open downgrade principle:

*Open downgrade principle:* If I have a reason to believe p, and substantive evidence that not-p arises, I now ought reduce my belief in p to some degree

The principle seems intuitively true. It clearly applies to testimony in empirical cases. Entitlements are supposed to model certain common sense intuitions of how knowledge gathering is supposed to work. If the downgrade principle were false, we would lose much everyday reasoning. Consider these two cases:

Case A: It seems to me that there is smoke in the distance, but somebody standing next to me disagrees.

Case B: I recall having given an assignment out to my students yesterday, but my students disagree.

Imagine how I would reason if the open downgrade principle were false. Case B is the easiest. If open downgrading were false, I could treat my memory as providing a settled belief. In that case, I could surely conclude that my student have all either misremembered or lied. This seems obviously wrongheaded; in Case B, I clearly ought to have some reason to suspect my memory. Case A is a little more difficult. It is plainest if we multiply the testimony. Imagine that hundreds of people claim there was no smoke in the distance. If open downgrading were false, I could discard each individual piece of testimony on the basis of its being contrary to my settled belief. But surely I ought not - even if the impact of each individual testimony is low, surely in the aggregate I ought to have some doubt against my vision. This seems to be a paradigmatic case for defeating entitlements.

Any entitlement theory which disallows open downgrading in general will fail to capture basic intuitions about how knowledge-acquisition and empirical reasoning is supposed to go. In order to capture basic intuitions about cognition, entitlements ought to be moderately vulnerable to defeat. They shouldn't be too vulnerable, since entitlement theory is supposed to provide a relief from endless calls to justification. For example, it shouldn't be such that another's merely pointing out that one doesn't possessing a proof for  $p$  constitutes a defeat for one's belief that  $p$ . The mere question, "How do you know your eyes work?" shouldn't, in and of itself, defeat one's entitlements. If, however, one receives good evidence that one might have ingested a hallucinogen, or that one might actually be a brain in a vat, then one's entitlements ought to be partially or fully defeated. A reasonable entitlement theory should provide a relief from trivial challenges, but demand that an agent respond to adequate challenges, or face defeat. But where should that bar for adequacy be? If we set the bar for adequacy too high, then we lose the defeasibility that helps make entitlement so reasonable.

I cannot hope to specify the criteria for adequate challenges here. Developing a complete account of adequacy seems to depend on committing to and developing a particular theory of entitlement. But it seems, *prima facie*, that any reasonable entitlement theory should grant that a contrary report from a source of comparable epistemic status must constitute an adequate challenge. Entitlement theories are built to model common-sense intuitions about how knowledge works. In the empirical realm, common sense clearly indicates that the presumptions of self-trust are defeasible through social pathways. We check our memories against the memories of others, the sensitivity of our

tastebuds against the reports of others. It seems, *prima facie*, that any reasonable entitlement theory ought to allow testimony and disagreement to defeat, at least partially, my entitlements to self-trust. I will assume that the open downgrading principle is true in general. The burden of proof, therefore, is on those who seek an exception to the principle in the moral sphere.

### **Simple transmission**

The most plausible path to blocking the open downgrading principle is the path of *dismissal*. Here, we grant that a peer might affect one's beliefs when one has no judgment of one's own on the matter. But once one decides for oneself what one thinks on the matter, peer testimony loses all its weight. Thus, the defender of moral self-sufficiency might grant that testimony can provide substantive evidence, but claim that *acquiring a judgment of my own defuses the evidential weight of contrary testimony*. The difference between the trumping argument considered in Chapter 2 and this new dismissal argument is that the trumping argument maintains that the weight of contrary testimony is featherweight but constant, while the dismissal argument allows that contrary testimony is initially substantive, but that its evidential weight evaporates for some reason when I acquire a judgment of my own.

Why might acquiring my own judgment defuse the weight of contrary testimony? On first blush, dismissal seems implausible. There seems to be *continuity* in the reasons one has to believe one's interlocutor before and after one has decided for oneself.

Suppose that an interlocutor's believing not-p provides me with a reason to believe not-p when I have no formed judgment about the matter. I must have some reasons for taking my interlocutor to be reliable. These reasons for trusting my interlocutor cannot emerge from considerations directly pertaining to not-p, because I have no direct considerations to believe not-p. The reasons for trusting must, then, emerge from general considerations about the reliability of the interlocutor - about their status as an epistemic peer. These reasons might include their being well educated or sensitive, or their having a history of accuracy in the relevant domain. The fact that I later discover my own reasons that p seems to be entirely unrelated to these general considerations for reliability. The reasons I have for trusting testimony have to do with the moral reliability of my interlocutor and nothing to do with particulars of the situation at hand, while the reasons I have for trusting my own judgment have to do with details about the situation at hand and my own moral beliefs, and little to do with my interlocutor. The two do not bear directly on each other, and so the process of forming my own judgment should have no bearing on my reasons for trusting my interlocutor. My trust in my interlocutor should be continuous between the situation of having no belief of my own, and the situation of disagreement. Thus, after I form my own judgment that p, the considerations for trusting my interlocutor, and thus testimonial reasons in favor of not-p, should still be active. This is not to say that I don't have plenty of new reasons to believe p; it is simply to say that whatever reasons I had to trust my interlocutor that not-p in the first place are still active.

But my continuity argument ignores one very crucial interaction between peerhood and making up my own mind about the matter. The argument depends on

presuming that there is no direct interaction between my trust in my interlocutor's claim that not-p and my judgment that p. But there is one very important interaction: through the very proposition itself. The very fact that we disagree over p is a reason to reduce my trust in another person. After all, one of our primary tools in evaluating another's reliability is their agreement with a wide body of conclusions we take to be true. If agreement establishes reliability, then disagreement ought establish unreliability.

I grant that there is a direct path to downgrading through disagreement, but this direct path isn't enough to rescue moral self-sufficiency. This is because, though disagreement does give us a reason to downgrade, the downgrading is bilateral. Suppose we accept the following principle:

Simple transmission principle: The fact that a rational source, S, disagrees with a trusted judgment, J, is a reason to downgrade confidence in S in proportion to one's confidence in J.

The simple transmission principle seems perfectly reasonable to me, but in a situation of disagreement, the principle applies to both sides of the disagreement. My interlocutor and I are both rational sources, and each us is in conflict with some judgment that we have an antecedent reason to trust. There is downgrading through disagreement, but the downgrading is bilateral. Even if the initial self-confidence is of a different degree from my trust in my interlocutor, this only shows that the resultant downgrading is asymmetric. This form of disagreement-based downgrading cannot eliminate my trust in

my interlocutor entirely, because simple transmission is bilateral, and thus the ensuing downgrading is bilateral.

Simple transmission is correct, but it does not block the core insight of the continuity argument. The original reasons to trust my interlocutor are continuous before I've made up my mind on the matter, and after. Once I make up my mind, I have acquired one new reason to downgrade my interlocutor - namely, that they disagree with me on this one issue - but I still have all the old reasons to trust. Since neither side is guaranteed, both sides must downgrade their trust in each other and in themselves.

### **Problems with self-discounting?**

Bilateral downgrading requires an attitude that might seem odd: it requires that I reflect on myself and render a judgment about my own reliability. I must treat myself as an authoritative source in such a way that I can stand back from that source and assign it a reliability rating. Bilateral downgrading requires that I treat my own judgments as having some qualified degree of reliability.

This may seem a strangely distant relationship to have with one's own capacity for judgment. The strangeness of this reflective, self-discounting stance is captured very well by David Enoch.<sup>40</sup> He argues that discounting one's own all-things-considered judgment results in a circularity. If you and I are both holding thermometers, and our thermometers disagree, says Enoch, it makes perfect sense for us to qualify our trust in

---

<sup>40</sup> Enoch (2011)



our thermometers. But this depends crucially on the fact that a thermometer is separate from our rational faculties, says Enoch. One cannot separate oneself from one's all-things-considered judgment; one cannot treat oneself as a mere truthometer. One cannot make a final all-things-considered judgment then discount one's confidence in it by some amount, because that latter, post-discounted judgment *is* one's all-things-considered judgment. Self-discounting isn't possible, says Enoch, because of the ineradicability of the first-person stance.

Enoch's argument seems to me quite plausible, but his argument applies only to discounting one's final, all-things-considered judgments. One can come to distrust a sub-faculty or particular process without running aground of the circularity - we do it all the time, and are surely reasonable in doing so. One can, for example, come to distrust one's memory - by, for example, discovering that one's memory for faces is quite poor.

To the degree that one's moral judgment depends on a subset of one's general reasoning abilities, then one can stand back from them and coherently engage in substantive self-distrust. It is no more paradoxical than standing back and using my cognition in general to discover evidence that my mathematical ability is poor, or that my ability to judge the character of people was poor. What's important here is that my general cognitive abilities are distinct from my domain-specific abilities. That is, my confidence in, for example, my mathematical judgments will have two parts: first, the output of my mathematical process, and second, my evaluation of my own ability to perform the mathematical process. A student that knows themselves to be very shaky mathematically can perform the mathematical calculation to the best of their ability, but

have reflective doubt in that calculation. The only thing that Enoch's argument forbids us from doing is discounting our all-things-considered judgment. The output of my mathematical operations is not an all-things-considered judgment, for it leaves out reflective information of the reliability of my mathematical abilities.

Self-discounting can happen for any subset of reasoning, but it is particularly pertinent for sealed cognitive faculties. One's memory and raw vision are, in an important sense, particularly like thermometers - they are faculties quite distinct from one's general cognitive faculties, and not entirely transparent to introspection. Since we cannot get inside them introspectively and check the steps and proper functioning of the mechanisms, we must rely on external data for reliability. We must look to their outputs, and check their outputs against the outputs of other purportedly reliable sources. Raw moral intuition, to the degree that it emerges without introspection and explicit reasoning, is sensitive to self-discounting for these same reasons.

Self-discounting is possible precisely because our reasoning process is divisible, and some of those sub-divisions are particularly vulnerable to self-discounting because they are sealed to introspection. Enoch-style considerations provide no viable pathway to moral self-sufficiency. In fact, the seeming reliance of most moral reasoning on raw moral intuitions make moral reasoning more susceptible to downgrading than other, more introspectively transparent cognitive processes.

## **Part II. Introducing Thomas Kelly**

Thomas Kelly presents, in "The Epistemic Significance of Disagreement," what I believe to be the best contemporary argument that disagreement in and of itself ought have no effect. His argument presents a sophisticated block to the open downgrading principle. His argument is strictly an analysis of how we ought to reason about disagreements; it makes no special reference to the epistemic status of the positions held on either side. In fact, I will claim that Thomas Kelly's analysis suppresses crucial epistemic considerations.

Kelly has two arguments for the epistemic insignificance of disagreement. The first argument is that we only rate others as peers based on agreement; for once we disagree with somebody on some topic, we ought to stop regarding them as a peer on that topic. Thus we will never be in the situation where we disagree with a peer. The second argument is that we ought take testimony to be weighty, but only as a proxy for the evidence itself. But once I see the evidence and reasons for myself, the testimony has no further weight in and of itself. Once I've settled the matter for myself, contrary testimony can only give me reason to reduce my trust in the testifier. I will discuss each of these arguments at length momentarily.

Kelly's arguments concern the nature of disagreement in general, but they have a particular application to moral disagreement. Kelly's arguments depend on there being epistemic parity between the disagreeing parties; disagreement only doesn't count when the agent has access to the all the same evidence and reasons as my interlocutor. But, in most empirical disagreements, I can easily suppose that my interlocutor might have

access to evidence and reasons that I don't. As a result, most empirical disagreements fall outside of the scope of Kelly's conclusions. But the paradigmatically problematic moral disagreements are the ones where both parties have access to the same evidence.

Distinctively *moral* disagreements occur when we agree on the straightforward facts of the matter, and only disagree in our moral assessment. Kelly's conclusions may apply to a few rare empirical cases, but seem to apply to all the cases I take to be paradigmatic of troubling moral disagreements, for precisely the reasons that I take them to be paradigmatic - that they are long-standing disagreements between peers. Kelly himself explicitly claims that his arguments apply to such paradigmatic cases of moral disagreement, and takes the dismissal of such moral disagreements to be one of the major consequences of his argument.

For the remainder of this chapter, I will take up Kelly as the best representative of moral self-sufficiency on epistemic grounds. I will argue that Kelly's argument leaves out a crucial factor of most moral disagreements: the fallibility and defeasibility of the involved faculties. Kelly's arguments succeed when our judgments are non-defeasible - perhaps, for example, short logical proofs - but fail when our judgments depend on defeasible faculties. Thus, they may succeed for those moral judgments that are grounded in non-defeasible proofs - but they will fail in any cases where an agent's confidence in her moral judgments depends on an entitlement to a loose judgment, including any reliance on moral intuition.

### **Kelly's conclusion, and its scope**

Kelly's discussion of disagreement focuses on the question of how we ought to weight *contrary testimony*.<sup>41</sup> He grants that testimony has weight when an agent has no judgment of her own, but claims that *testimony contrary to an agent's considered judgment has no weight* in a situation of epistemic parity. I'd like to begin by discussing the scope of his conclusion.

It is easy to think, contra Kelly, that contrary testimony from a rational, reliable person ought to have some sort of weight. As I've argued in previous chapters, this position of weighting contrary testimony arises naturally from two basic epistemic commitments. First, we think testimony can often furnish reasons to believe. We often rely on the observations of other people - eyewitnesses in jury trials, fellow scientists publishing experimental data. We also rely on the judgments of experts - our doctor, our car mechanic. If contrary testimony provides an undismissed good reason to believe the contrary of our own judgment, then, *prima facie*, it seems it should reduce our confidence in our own judgment.

Kelly's position navigates between these commitments. He allows that disagreement can provide reasons to believe in many circumstances, but in those cases it is not the mere fact of disagreement that's doing the work. Rather, what's doing the work is the fact that one's disagreeing interlocutor has access to information or abilities one does not. Once we pare away all these other factors, and look at the impact of only the disagreement itself, claims Kelly, we will see that disagreement in and of itself can have

---

<sup>41</sup> Kelly (2006)

no effect. His claim is quite specific: he claims that *contrary testimony* has no weight in a situation of *epistemic parity*. Parties are in epistemic parity regarding p if and only if:

1. They are approximately as intelligent, well-educated, and rational in general.
2. They possess the same body of evidence as to p.

Individuals who satisfy both conditions relative to each other are considered each others' *epistemic peers* on the topic of p.

There are plenty of cases where do we take contrary testimony, but Kelly explains them away as deriving from situations of epistemic non-parity. Thus, Kelly's claim is compatible with many everyday uses of testimony. One scientist may weight the contrary claims of another scientist, because the other scientist has access to evidence that that the first does not - different observational data from a different clinical trial, perhaps. The claims that my math professor makes about the wrongness of my proof has weight because my math professor is better educated than me in the relevant domain. The controversial situations, says Kelly, are those where "I have a firmly-held belief, [and] there are some who disagree with me whose judgment cannot be simply discounted by appeal to considerations of intelligence, thoughtfulness, or ignorance of the relevant data."<sup>42</sup>

Kelly's description of epistemic parity fits precisely my paradigmatically important cases of moral disagreement. Thus, Kelly and I are in precise opposition on this

---

<sup>42</sup> Kelly (2006), p 2

core set of cases. Claims Kelly, "once I have thoroughly scrutinized the available evidence and arguments that bear on some question, the mere fact that an epistemic peer strongly disagrees with me about how that question should be answered does not itself tend to undermine the rationality of my continuing to believe as I do."<sup>43</sup> Continuing to "confidently retain my original view in the face of such disagreement" is not irrational, says Kelly, and "might very well be the uniquely reasonable response."<sup>44</sup> Thus, Kelly can explain away many of the everyday cases of the seeming weightiness disagreement as situations of epistemic disparity; his arguments target quite specifically the weight of another agent's *judgment to the contrary* from an identical set of evidence.

Kelly intends his conclusion to cover moral disagreements. One of his three basic examples of dismissable disagreement is the issue of whether "Truman's decision [to drop the atomic bomb] was morally justified."<sup>45</sup> Moral cases like the Truman case are apt targets for Kelly's arguments, in virtue of his conditions for epistemic parity. His argument only applies when we believe ourselves to have all the same evidence as our interlocutor, and take ourselves to have fully scrutinized that evidence. These strict criteria for epistemic peerhood will rarely apply in everyday cases of disagreement over empirical matters of fact, but such disagreement between epistemic peers is the paradigmatic situation of the epistemically significant moral disagreement. I take it that we are supposed to image a dispute between two contemporary interlocutors about Truman's historical decision. Since the Truman case is historical, and thoroughly public,

---

<sup>43</sup> Kelly (2006), p 4

<sup>44</sup> Kelly (2006), p 5

<sup>45</sup> Kelly (2006), p 1

then most ordinary moral disputants would be in an identical epistemic relationship to the facts of the matter. The strictly *moral* disagreement occurs once we've reached agreement about all the pertinent historical facts.<sup>46</sup> Kelly's conditions of peerhood are precisely the conditions of moral disagreement raised in favor of moral skepticism. The reason why intractable moral disagreement are supposedly so epistemically remarkable is that they survive even after we've agreed on the description of the state of affairs in question.

Thus, I take Kelly to be making an exemplary argument for the dismissal of contrary testimony on epistemic grounds.<sup>47</sup> His argument is not targeted at moral disagreements uniquely; rather, they are targeted at the category of disagreements in situations of epistemic parity. If Kelly is right, his claims will fit quite neatly with the intuitions that dismissal is rarely warranted in empirical matters, but are par for course in moral matters - since epistemic parity is so rare in empirical disagreement, and so common in moral disagreement. Thus, Kelly's model neatly fit our intuitions of the disanalogy between empirical disagreement and moral disagreement, without invoking any special properties of moral disagreements. It's simply an empirical feature of moral disagreements that they often occur in situations of epistemic parity. Furthermore, Kelly's arguments fit many of the intuitions about which cases moral self-sufficiency applies to. Most of my opponents grant that moral disagreement matters in various side-cases - for example, in moral education, when I am young and inexperienced, and you are wise in

---

<sup>46</sup> Even we consider moral intuitions to be a form of evidence, that, too, can be shared - I may know that my friend sees it as entirely obvious that mass murder is never justified, and she knows that I see it as undeniable that killing is often justified in the defense of one's country.

<sup>47</sup> Kelly's argument strictly concerns the question of whether contrary testimony may be dismissed on epistemic grounds. The question of whether contrary testimony may be dismissed on moral grounds - say, as a violation of personal integrity or moral autonomy - is a separate issue, which I take up in Chapter 4.



the ways of the world; or, for example, in taking advice on how to deal with an angry girlfriend, from somebody who has more knowledge and experience about her inner workings. But for the paradigmatic moral disagreements, the ones where we seem to have access to all the same facts and reasons, the intuitions seems to be that *here we ought to make up our minds for ourselves*. This split in intuitions is again captured precisely by Kelly's criterion of epistemic parity.

### **Part III: Kelly and the Argument from the Evaluation of Rationality**

Kelly has two significant arguments for his conclusion, which I will consider separately. His first argument proceeds from an analysis of the grounds for judging others to be peers. Before I present Kelly's actual argument here, I would like to discuss a more primitive version, which I'll call the direct argument. Though the direct argument is obviously flawed, it will be useful to make these flaws as clear as possible, for these flaws recur in subtler form in Kelly's sophisticated version.

The direct argument claims that, if I've judged that p, and you offer testimony that not-p, then I can dismiss your testimony precisely because you believe not-p when really, it's p. Your belief that not-p is direct evidence that you've failed to reason well about p. Suppose I've read some arguments and made up my mind that it's wrong to eat animals. My friend, who is as philosophically astute as me, claims he thinks that it's right to eat animals. Normally I would give some weight to my friend's advice, but here, I already know it's wrong to eat animals, and so I may, on the basis of this knowledge, conclude

that my friend is wrong in this case. And, since I know he's wrong, I should accord his contrary testimony no weight. In the direct argument, *my assurance that p underwrites my dismissal of any interlocutor that believes not-p.*

The direct argument violates the open downgrading principle and bilateral downgrading. The direct argument depends on an agent's treating themselves as infallible. Without the presumption of infallibility, direct dismissal would be circular - I'd be settling the question of whether or not I was reliable in judging p by referring to my judgment that p. But the presumption of infallibility is unreasonable under entitlement theory, as we have already seen. The direct argument depends on, at best, a terrible sort of rational egotism, and, at worst, solipsism.

Kelly presents an improved version of the direct argument that lacks this commitment to rational egotism. He freely grants that we are epistemically fallible and that we are each but one rational being among many. But, argues Kelly:

Whether we find the possibility of disagreement intellectually threatening, I suggest, will and should ultimately depend on our considered judgment about *how rational* the merely possible dissenters might be in so dissenting. And our assessment of whether rational dissent is possible with respect to some question (or our assessment of the extent to which such dissent might be rational) will depend in turn on our assessment of the strength of the evidence and arguments that might be put forward on behalf of such dissent. But if this is correct, then the extent to which merely possible dissent should be seen as intellectually threatening effectively reduces to questions about the strength of the reasons that might be put forward on behalf of such dissent. ... The role of disagreement, whether possible or actual, ultimately proves superfluous or inessential...<sup>48 49</sup>

---

<sup>48</sup> Kelly (2006), pp 18-19

<sup>49</sup> Kelly talks about the possibility of disagreement here, rather than the existence of actual disagreement, because he thinks that possible disagreement is more important than the existence of actual disagreement. This is because actual disagreement is vulnerable to non-rational forms of resolution - for example, a tyrant may remove disagreement by having all disagreeers put to the death. I do not believe the distinction has any importance for my discussion.

Provided that we are in the situation of epistemic parity, Kelly argues that our evaluation of the epistemic weight of the disagreement of an interlocutor will depend on our evaluation of the rationality of that interlocutor. This in turn depends on our evaluation of the strength of their reasons. If their reasons seem poor, then we will dismiss the interlocutor. And, in making a judgment contrary to theirs, I'm judging that their reasons are, in fact, poor. In short, once I settle for myself that p, I cannot judge that somebody is, or could be, rationally dissenting with me on p - for their very dissent shows that they are, to the best of my own ability to judge, not rational on the issue. I will call this the *argument from the evaluation of rationality*, or the argument from evaluation, for short.

Kelly's argument from evaluation, though similar in spirit to the direct argument, avoids the latter's commitment to rational egotism. It is predicated not on an agent's taking themselves to be infallible, but on an agent's epistemic limitations. We cannot know automatically whether the people around us are rational and reliable; we need grounds for evaluating them. The only way to judge the reliability of others is by their outputs, and the only way to judge those outputs is in terms of what I take to be correct.<sup>50</sup> We trust others based on their agreement with us. Thus, in situations of disagreement, the direction of downgrading must always be in the favor of our judgments over interlocutor testimony - not because we are infallible, but because the only way we can assess interlocutor peerhood is through agreement with what we take to be the case, which, in turn, we can only access through our judgments.<sup>51</sup> In other words: since others' agreement with one's own judgments is the *source* of one's trust in others, then we would be

---

<sup>50</sup> Kelly's argument here seems to follow the lines of Davidson's comments on radical translation.

<sup>51</sup> Kelly's argument here has a passing resemblance to Donald Davidson's discussions of radical translation.

unreasonable in trusting another above and beyond their agreement with us. Our epistemic limitations pervade every aspect of our reasoning, including our ability to evaluate the rationality of others. For Kelly, it is not that we know ahead of time that we are infallible or that they are irrational; it's that our confidence in their rationality can never exceed our confidence in our own.

This argument I take to be a significant improvement on the direct argument; it falls, however, to a more sophisticated version of the same problem.

### **The Unity of Rationality**

Kelly's argument for dismissal from the evaluation of rationality is incorrect; it depends on a *radically disunified conception of rationality*. Suppose that Tim takes John to be an epistemic peer, such that John's believing something is sufficient to give Tim some reason to believe it. Suppose that Tim has a settled judgment that p. Then, he meets John, and John claims the contrary. If Kelly's argument is to work, then the mere fact of disagreement over p alone must be sufficient grounds for Tim to think John *entirely unreliable* on the topic of p. Kelly's argument depends on the truth of the following principle:

*Disagreement Dismissal Principle (DDP)*: An interlocutor's believing against one's settled judgment that p is sufficient grounds to dismiss that interlocutor as entirely unreliable on the topic of p.

On what grounds could DDP rest? It might rest on the claim that agreement was the sole grounds for taking the interlocutor to be reliable in the first place.

*Agreement as Sole Grounds Principle (ASGP):* The only relevant evidence of an interlocutor's reliability on the issue of p is their agreement with one's independent judgment on the issue of p.

If the only reason to trust an interlocutor on the topic of p is agreement *on the particular topic of p itself*, then disagreement would indeed rob one of all of one's grounds for trusting that interlocutor. Thus, ASGP supports DDP. But ASGP is obviously absurd. As one consequence of ASGP, I could never gather new beliefs from testimony. If I had no belief on whether p or not-p, I could never have the grounds to trust anybody's testimony on the matter.

But we do gather new beliefs from testimony all the time. It follows, then, that ASGP is false, and that *there must be more general ways to establish an agent's rationality on p than their agreement with me that p*. This latter claim is quite everyday. I trust my mechanic's claim that it's the fuel injection system that's the problem, not because he agrees with my judgment about the fuel injection system, but because he's fixed my car in the past, and he seems to know his way around an engine, and he has many satisfied customers. It's a good thing that I can do this, because I know nothing about fuel injections whatsoever. I trust a restaurant reviewer's recommendation based on

the degree to which they seem to express sensitivity and passion, the sort of details they pick up on, and their past reliability on other restaurants. If I had to agree with the restaurant reviewer on a particular restaurant before I could trust his recommendation of that restaurant, then the whole enterprise of recommending restaurants would be useless.

AGSP rests on a presupposition of the radical disunity of rationality. It supposes that that we judge others to be peers proposition by proposition, and that success on one proposition confers no probability of success on a related proposition. But AGSP must be wrong; in order for ordinary social knowledge gathering to work, there must be some other means to assess reliability on a particular topic than agreement on that very topic, and that very topic alone. My argument against AGSP here is, in a way, an extension of the old puzzle about teaching. If the only way to be able to judge that a teacher is a good teacher of a particular domain is to already have domain-specific knowledge, then finding a good teacher to start learning about a new domain would be impossible. Since we can find good teachers for new domains, then there must be other ways to judge good teachers besides domain-specific knowledge. There must be general signs of intelligence, wisdom, and ability that we can recognize that allow us to form reasonable judgments about who is likely to be, say, a good teacher of pottery, when we presently know nothing about pottery.

But if we drop ASGP, then Kelly's argument from the evaluation of rationality fails. Once we drop the commitment to disunity, we will see that we are allowed many forms of evidence that somebody is reliable on the topic of  $p$ , of which agreement on  $p$  itself is merely one. For example, let us say that Fred contests my claim that the proper

etymology of "knight errant" is "knight on an errand." I have a good reason to believe that my etymology is correct - I remember reading it somewhere, and it makes sense to me. And so I have good reason to disbelieve Fred. But I know many other things about Fred - he's smart, he's well-read, he's got a good memory, and he's often right about neat historical facts. So I have all these reasons to think that he's right, and I'm wrong. It is very likely that this trust in Fred is not enough to settle the matter conclusively against my own judgment, but it's surely worth something. In other words, if we drop ASGP, then the following must be true:

*Other Grounds Principle (OGP)*: There is relevant evidence for an interlocutor's reliability on the topic of p, other than agreement with one's independent judgment that p.

But the argument from evaluation fails to completely eliminate contrary testimony, once we grant OGP.

The only way to make DDP true in the face of the eminent truth of OGP is by subscribing to the following principle:

*Disagreement Trumping Principle (DTP)*: An interlocutor's disagreement with my settled judgment on the topic of p yields evidence that the interlocutor is entirely unreliable, and that evidence trumps and eliminates all other evidence of the interlocutor's reliability.

If DTP were true, then OGP wouldn't matter. The general reasons for trusting Fred would be trumped by the specific reason that he believes the wrong thing. But DTP fails for the same reason that the direct argument fails: it presumes the infallibility of my own judgment. If we grant the possibility of our own fallibility, then bilateral downgrading applies once again. All the specific reasons I have to believe in my etymology are also reasons to think that Fred is unreliable here, and all the reasons I have to think that Fred is reliable are evidence that I am unreliable. In short, if I have evidence that p, and my interlocutor disagrees, my evidence that p would be *only one reason to dismiss them against many reasons to trust them*. It may be that my reason to believe is rather strong, and my reason to trust them is rather weak - I just read the etymology in a book the other day, and Fred's kind of absentminded and prone to invention. Or it might be the reverse - my memory for this sort of thing is only mediocre, but Fred's memory for trivia is spectacular. In either case, the principle of bilateral downgrading applies, and so complete dismissal and elimination is impossible. After disagreement, we may accord our interlocutor less epistemic weight, but we still accord them some epistemic weight.

In order to support the crucial DDP, Kelly is driven either to subscribe to a radically disunified view of rationality, or to the same problematic egotism that made the direct argument unpalatable. Since both are untenable, we must reject DDP - and without DDP, we cannot eliminate contrary testimony via the argument from the evaluation of rationality.



## Diagnosing Disunity

Why would somebody, then, subscribe to DDP in the first place? Perhaps it is because the crucial AGSP is very similar to some other, very plausible principles. Take, for instance:

*Contraries Undermine Principle (CUP):* Disagreement with my best judgment on the issue of p is good evidence against an agent's rational reliability on the issue of p.

CUP is quite sensible, but it won't yield elimination. CUP yields that disagreement is *a* reason to think that a disagreeing interlocutor is unreliable. But it doesn't say disagreement and agreement are the *only* reasons bearing on the reliability of the interlocutor on this topic. CUP leaves the door open to bilateral downgrading. CUP is correct, and may lead to a reduction of confidence in my interlocutor, but it will not yield complete elimination.

Another principle, very subtly different from CUP, is:

*Correctness Principle (CP):* The only relevant evidence for me of an agent's rational reliability on the issue of p is their correctness about p.

This principle seems almost tautological. But it will not provide us with elimination, because of the epistemic space between "my best judgment about p", and "correctness about p". In short, in order to apply CP, I would need direct, antecedent access to the

truth of p - which I don't have. In order to eliminate an agent entirely using CP, I would need to be entirely confident about my judgment that p. But I can't be - at the very least, because I have a piece of undismissed contrary evidence from a reliable resource. CP requires a god's eye view.

Though both these principles are plausible, neither will support Kelly's argument here. The crucial move for Kelly is the claim that our evaluation of the rationality of a dissenter on a given topic depends *solely* on our evaluation of their evidence and reasons for their belief on that topic. I have argued that this claim depends on an excessively limited view of what may contribute to an evaluation of an interlocutor's rationality. Once we see that we must allow more sources of evidence for the rationality of others, the argument from evaluation falls.

#### **Part IV: Kelly on Independent Reasoning and Proxies**

Kelly's second argument for dismissal is based on the importance of independent reasoning. The mere fact of disagreement ought not have any weight, Kelly argues, over and above any convincing force that the arguments advanced by the other side have. Once again, his argument is a sophisticated version of a more primitive argument. I will begin with the primitive version.

## Rejecting all testimony

The primitive claim is simply that *we ought never simply take another's word for something*. One might think that other people can point us to evidence we haven't seen, or introduce us to reasoning we haven't thought through ourselves, but may never simply tell us what to believe. To make this claim, one must subscribe to the following principle:

*No Testimony Principle* (NTP): The mere fact that a trustworthy source testifies that p cannot give one a reason to believe p.

NTP implies that an interlocutor may only point out new evidence and present argument for me to judge, but that I must evaluate the evidence and think through the arguments myself.

But NTP is utterly absurd. To summarize various considerations from earlier in this dissertation: NTP makes all forms of testimony unusable, but testimony is vital to broad swaths of human activity. Very convincing work has already been done on this subject. Annette Baier has argued that any form of cooperative human activity requires trust (1986, 1992). John Hardwig (2005) argues effectively that without trust in testimony, much of our normal epistemic life would be impossible. Legal trials depend on using the observations of others - when a jury hears eyewitness testimony, they are taking another person's word for evidence that they cannot observe for themselves. Modern scientific work depends essentially on using the observations of others, argues Hardwig. A scientist can't observe every experimental result herself; she has to accept the

experimental reports of other scientists in other laboratories. Furthermore, she can't work out every result from all related fields of science and mathematics myself. She simply don't have time to collect the data herself, perform the statistical analysis herself, etc. etc. She has to trust not only the observations, but the *judgments* of other scientists.

NTP is too a stringent a requirement for human lives. If NTP were true, for example, I could not ask somebody what time it was and take their word for it; I would have to see the watch for myself. There is simply far too much evidence for a single person to sort through on her own. Science, and much of daily life, requires the delegation of epistemic tasks.

### **Double Counting and Faculties**

NTP is obviously flawed; it forbids us from ever using testimony in any sensible way. Any successful argument claiming the importance of independent reasoning must allow for such everyday uses of testimony. Defenders of moral self-sufficiency must allow for some unavoidable uses of testimony, while dismissing testimony in situations of peer disagreement.

Kelly's presents a novel model of the use of testimony, which aims to avoid the flaws of NTP, while stop arguing for essential epistemic self-sufficiency. Kelly claims that we use testimony only as a *proxy* for the testifiers' actual grounds. That is, if somebody reliable believes *p*, and I have no idea what their grounds are, I may treat their belief that *p* as indicating the presence of good reasons for *p*, without actually knowing

what those reasons are. For instance - if you run in shouting that there's an elephant running around loose outside of the circus, then I may use your claim to justify a belief that there is such an elephant. But it's not your *belief* in the elephant that's doing the work; it's your presumed grounds. I assume that you either saw the elephant, or had sufficient grounds to infer that there was an elephant. I use those presumed grounds to justify my belief - I can't say what those grounds are exactly, but I have reason to think they're good grounds. After all, you're a rational person, and you believe it, so you probably have good grounds.

But a proxy is merely a stand-in for the evidence. Once I go into the circus tent and come into direct, unmediated contact with the evidence, your testimony plays no further role. I cannot add your testimony as to the elephant to my seeing the elephant and be doubly sure. Rather, my seeing the elephant for myself *replaces* your testimonial proxy and renders the testimony epistemically unimportant. I no longer need a proxy for the evidence, since I have the evidence for myself.

Instead of NTP, Kelly subscribes to the following principle:

*No Mere Belief Principle* (NMBP): The mere fact that another rational being believes p, in and of itself, can never be a grounds for one's believing p.

NMBP is compatible with the use of testimony as evidence of other matters. The fact that you believe makes it likely that you have good grounds, and those implied grounds can serve as the grounds for my own belief. But once I've seen the actual grounds for your

belief and decided for myself that they're no good, then the mere fact that you disagree plays no further role.

Here's the argument in Kelly's terminology. Imagine that I believe H, on the basis of some set of non-testimonial evidence, E. Suppose that I discover that you also believe H, on the basis of the very same body of evidence. I shouldn't treat our agreement as *further confirmation of H* - that would be double-counting the evidence.

I am thus in the awkward position of treating your belief that H as a reason to believe that H, despite the fact that you do not treat this as an epistemically relevant consideration. Again, it might make sense for me to treat your belief in this way if I lacked access to your first-order evidence: in that case, your belief stands in as a sort of proxy for the evidence on which it is based... But when I do have access to your first-order evidence for H, and I continue to treat the belief that you have formed in response to that evidence as a further reason to believe that H, aren't I essentially engaged in a kind of double-counting with respect to the relevant evidence?<sup>52</sup>

When I list my evidence, E, for believing H, I may not enter *the fact that I myself believe H as additional evidence for H*. That would be double counting the evidence, for the belief that H, in some sense, is simply a recapitulation of E. This is the basic case of double-counting. Therefore, Kelly argues, if I were to use the fact that you believed H as a further reason for my believing H, that would be making precisely the same mistake. Since H is just a proxy for some set of evidence, E, and your believing H is just a proxy *for that same set of evidence*, E, then entering your belief in H as part of my body of evidence, E, would *make the same error of double-counting as in the basic case*.

There is something quite intuitive about this argument. Imagine, for example, that you and I just read the same history book. Let us suppose, for simplicity's sake, that we read the same physical copy. Suppose I were to cite some bit of trivia from it at a cocktail

---

<sup>52</sup> Kelly (2006), p 26

party, and you were to agree that the bit of trivia was, in fact, true. Furthermore, suppose we know that both of us claim this bit of trivia solely on the basis of having read it from the same copy of the same book. I shouldn't use your agreement as further evidence. Since our grounds are identical - that is, since our grounds is that this one particular history book says so - counting your agreement as further evidence would be basically counting the same book twice. My belief and your belief appear to be independent, but since they are based on the same evidence, they are actually not independent at all; increasing our confidence by taking both beliefs into account would be double-counting.

Thus, if Kelly's model is correct, we may use testimony in those cases when we have no access to the evidence ourselves, but once we do, the testimony loses all epistemic value. Using testimony and our own evaluation of the evidence would be double-counting. Thus, the model supports a sophisticated form of self-sufficiency, applicable to all cases of disagreement under epistemic parity.

### **Confirmation and agreement**

Kelly's double-counting argument makes a fundamental mistake. It presents an oversimplified view of the relationship between all-things-considered beliefs and first-order evidence. Kelly claims that my using *my own* judgment that H as a further reason for *my* believing H is double counting. This much I will grant. But he also claims that *my* using *your* judgment that H as a further reason for my believing that H is double counting

on the same grounds. But this is a mistake, for my relationship to my own judgment is quite different from my relation to your judgment.

When I take my own judgment that H as a further reason that H, I've added no piece of evidence. When I take *your* judgment that H as a further reason that H, I have added a further piece of evidence - namely, that your rational faculties are in agreement with mine. When I take the fact that we agree into consideration, I'm not doubling the evidence - I'm taking the agreement as further grounds to believe that both our rational faculties are functioning well in judging from the evidence. When I think I see a ship in the distance, and you chime in and say you see it too, this increases my confidence. But I'm not increasing my confidence by entering into my epistemic accounting the claim "There is a ship there" twice. I'm entering the claim, "I see a ship there", and the distinct claim, "Your eyes also see the same ship." There is added data here, for agreement is a form of confirmation of my own eye's reliability. That is, what I'm adding with the agreement is not the ship, again, but the fact that my visual system has been confirmed by another purportedly reliable source. This process of confirmation is nothing strange - it is, in fact, incredibly mundane. It happens when students in a math study group solves the same problem independently, then checks to see if they agree with each other. If they all got the same answer, this is good evidence that they did things right (though not certain evidence). The fact that their rational faculties agree is a reason to think they've all mastered the skill.

The double-counting argument would only work if an agent's believing in H reduced entirely to E, the evidence for H. But it doesn't. Belief in H depends both on the



evidence, and my faculties and skills for evaluating that evidence. It makes sense that I can't treat my belief that H as further evidence for H. I've used one body of evidence, and my own faculties, to arrive at H. The fact that I believe H adds no extra data to the reasons to believe H. But *your* believing H, though it may involve the same body of external evidence as my believing H, also introduces new information about the deployment and output of a distinct agent's faculties and skills. When I take into account the fact the you agree with me that H, on the basis of H, what I'm taking from you is not E again, *but the further fact that your faculties are functioning similarly to mine in this case.*

Imagine that you and I read the same book in college, ten years ago. I claim a bit of trivia from it, and you agree. This agreement does in fact add confirmatory weight. The original, external evidence is precisely the same between us, but the agreement adds a degree of confirmation to the proper functioning of our faculty of memory. We're not adding in the information that there's another book that also claims this fact; we're adding that we both remember the same thing about the same book, and so are less likely to be in error about what we remember.

In precisely the same way, the fact that somebody disagrees with my evaluation of the same evidence is not some strange double-existence of the evidence in opposite forms - it's the same evidence, and two different judgments based on the evidence. Kelly's mistake, in thinking that judgments are a proxy for evidence, is to think that there's no distance between the evidence and the judgment. It's to think that our judgment is simply a transparent window to the evidence. But there is distance. In most everyday judgments,

we deploy some fallible faculty. This is why we use disagreement and agreement as checks on functioning. If two of us use the same data and perform the same calculation, and we get the same result, this is confirmatory. If we get different results, this is evidence that something has gone wrong with one of our calculations. It may be that it's more likely to have gone wrong with my calculation rather than yours or vice versa, depending on our degree of ability and our present degree of drunkenness and so forth. But the mere fact that a judgment is *mine* doesn't guarantee it. At least not when the judgment relies on defeasible faculties.

Now, there is an imaginable response on Kelly's behalf: that what I've shown is not that the double-counting argument is wrong, but that two agents are never in true epistemic parity. They always have access to different data; I have access to the data of my eyes, and you have access to the data of your eyes, I have access to my own reasoning faculties, and you have access to yours, etc. etc. In that case, Kelly's claim is true but vacuously fulfilled; there would be no disagreements that could fit *that* criterion of epistemic parity, because no two agents could ever truly be in a relationship of true parity. In any case, I think the central insight of my argument remains the same: in any possible disagreement, there will be some epistemic information that will alter the all-things-considered judgments of the involved parties. The information is reflective information about the agents' own well-functioning. Kelly's insight about double-counting might explain why I don't treat the belief of others as equally important as mine, but it cannot render disagreement between peers entirely insignificant.

## Proofs and Entitlement

The prominent role of fallible, defeasible faculties in the preceding discussion also points out exactly those conditions under which Kelly's argument would, in fact, be plausible. In order for the Kelly double-counting argument to eliminate disagreement entirely, my judgment  $p$  must be sealed against the possibility of faculty malfunction. Let's start with a degenerate case, for simplicity's sake. If, for example, my evidence was  $p$ , and my judgment was that, in fact,  $p$ , then there's no work done by any fallible faculties of judgment, and no room for error. If I counted my judgment that  $p$  based on the evidence that  $p$ , and then added to it your judgment that  $p$  based on the evidence that  $p$ , I would, indeed, be double-counting  $p$ .

Simple entailments will work the same way. Let us say that the evidence,  $E$ , consists of  $p$  &  $q$ . I judge based on the evidence that, in fact,  $p$ . If you judge on the very same evidence that  $p$  then it seems like my using the fact that your judgment agrees with mine would likely be double-counting the evidence. This is because your agreeing judgment adds no information to my judgment.

This is, perhaps, why Kelly's argument seems so plausible: it *is* correct for special cases where there is no exercise of a defeasible faculty. Take, for instance, any situation where the relationship between evidence and judgment is a very simple and straightforward proof. Let us say that the evidence consists of the fact that  $p$ , and the fact that, if  $p$ , then  $q$ . I judge that  $q$ . You refuse to judge that  $q$ , and say that, even though you agree with all the evidence, it just doesn't seem to you that  $q$  follows. In this case, I *can*

dismiss you because, in simple proof case, there is no possibility of error. If the judgment involves the use of no faculties whatsoever, or only uses non-defeasible faculties, I can't stand back from my own judgments in order to downgrade my degree of self-trust. There's simply no room to do any downgrading. We can't do bilateral downgrading, because there's no aspect of myself vulnerable to downgrading - so it's got to be my confidence in my interlocutor that gives. Kelly's argument does work in cases where there's no reliance on any defeasible faculty.

But these are rather specialized, and rare cases. (Some even hold that there are no non-defeasible judgments.) In most everyday judgments, there is quite a large amount of room between the evidence and the judgment. Visual perception obviously involves the use of a defeasible faculty. Complex evaluations like, "This table is sturdy," and "He's not very trustworthy," surely invoke defeasible faculties of judgment. And unsecured moral judgments involve a defeasible faculty, by definition. Kelly's argument fails to apply to most non-trivial reasoning, because it fails to take into consideration the fallibility of human cognition.

### **Explaining the original case**

Why, then, was Kelly's original argument so convincing? Recall the case: I hear your testimony that there is an elephant and then I see the elephant for myself. It does, indeed, seem true that once I see the elephant for myself, another's testimony as to the existence of the elephant is rendered irrelevant. The epistemic force of the original

testimony seems to be entirely displaced by my own experiences and reasoning. Kelly naturally concludes that, as a general matter, getting the evidence for oneself neutralizes the relevant testimony.

I've argued that this neutralization cannot be complete. I now owe an explanation for our intuitions of neutralization and dismissal in cases like the one above. We are, I grant, right to think that agreeing testimony that there is an elephant doesn't really matter once I see for myself. But this is not, I argue, due to a universal trumping of testimony by direct evidence. Rather, it is a special property of certain cases where we already have a tremendous amount of confidence in the faculty involved. It's important that, in the elephant case, what's involved is visual perception of a nearby large objects. Visual perception of nearby objects is typically a heavily vetted, reliable system. For most of us, the visual perception of nearby objects is about as confirmed as a faculty could be.

I've claimed that agreement and disagreement from interlocutors doesn't give us a second entry of the evidence; rather, it provides a confirmation or disconfirmation that our faculties are properly functioning. *When the faculty is already well-confirmed, further confirmation provides a negligible change in confidence.* Given that I have a very high level of trust in my visual system, the confirmation provided by agreement between my brother and myself provides no, or a negligible, change in confidence.

I believe my reading to be more accurate to the phenomena of daily life than Kelly's, for two reasons. First, while agreement with an already well-confirmed faculty has little epistemic effect, agreement with others is part of what confirms the faculty in the first place. Part of what makes me so confident in my eyes is the constant, daily

agreement between what I see and what others see. Second, my explanation captures the asymmetrical worth of testimony in cases like this. If Kelly is right, then *neither testimonial agreement nor disagreement* will have any epistemic weight. But, intuitively, in situations like this, only agreement is neutralized. If I am standing next to my brother, and we agree, there seems to be no epistemic boost from the fact of agreement. But, if I am standing next to my brother, and he says he sees something, and I don't, I'll start to worry. If *many* people disagree with me about simple visual perceptions, I'll have plenty of reason to worry. The best explanation for this asymmetry is that the lack of epistemic effect from agreement is due to the fact that *there's no more room for any extra confirmation* of some systems (or at least that the system is so well confirmed that the effect of further confirmation is negligible). But there's plenty of room for disconfirmation of a high-confidence system.

It is true that, in the elephant case, the testimony does little work. But that's not because it's merely testimony, but because, in that particular case, there's no work for confirmation left to do. The visual system is already pretty much as well-confirmed as a perceptual system could be. To make this clear, simply re-imagine the case without my high degree of confidence in my own visual system. Suppose that I took a hallucinogen and spent a day hallucinating wildly. When I wake up in the morning, and I see a giant iguana in the room and begin to worry that I'm still in the throes of drug-induced hallucination until you say you see it too, then I've really learned something - that my eyes are working properly, and that there really is an iguana in the room.

#### **IV. Settling the Matter**

Both of Kelly's arguments fail for the same reason: they depend on an agent's treating herself as infallible. It is important to distinguish the point of infallibility to the logical point, that one cannot step back from one's all-things-considered judgment. In the all-things-considered sense of "judgment", one's judgment is authoritative of what one believes. Because how could one believe otherwise than what judges? But this logical point is distinct from Kelly's point. Let us distinguish two conclusions:

1. Logical Finality Principle: One's considered belief must reflect one's all-things-considered judgment.
2. Exclusive Self-Trust Principle: One's belief must reflect only the reasoning one does directly from non-testimonial reasons and evidence, and exclude any influence from reasoning done by others from reasons and evidence.

The Logical Finality Principle is the conclusion of Enoch's argument, and is satisfied by any instance of coherent human reasoning. It is an analysis of the notion of considered belief, and not a normative principle to guide our thinking. Insofar as we are forming our beliefs through reasoning, we are obeying the Logical Finality Principle. When I form a moral judgment, then discount this moral judgment because of the existence of

disagreement, my all-things-considered judgment is the product of these two considerations. I can discount the conclusions of a sub-faculty in forming my all-things-considered conclusion without violating the Logical Finality Principle. Thus, the Logical Finality Principle is obviously true, but it does not imply moral self-sufficiency.

Kelly argues for moral self-sufficiency via the Exclusive Self-Trust Principle, a much more radical claim. He argues that testimony, in and of itself, is specifically excluded from the list of possible evidence and reasons. Our use of testimony is merely a proxy for real evidence and reasons, and doesn't constitute a weighting of the testimony itself. In claiming that the testimony by itself cannot count, Kelly presupposes that I can take myself to be the final authority on epistemic grounds. When an interlocutor disagrees with my judgment, thinks Kelly, I may safely ignore the testimony, because it contributes nothing beyond the evidence and grounds I have already considered in making up my mind in the rest place. Kelly suppresses the contribution of disconfirmation; he treats an agent's judgment as the final arbiter in disputes about matters about the reliability output of my judgment, and the proper functioning of my judging abilities.

But this supposition cannot be sustained after a serious consideration of the epistemic status of unsecured moral judgments. Once I've confronted the evidence of shared human faculties, I must take others to be potentially authoritative as to the truth of various moral claims. Given that my mental faculties are fallible, and that my trust in them is defeasible, I must be open to downgrading my trust in my mental faculties. Once we see that testimony can be taken to impugn the reliability of my faculties, then Kelly's



argument can be seen as another version of that same mistake as the direct argument. Recall, from the discussion of the direct argument: if I believe p, and you claim the contrary, I cannot simply conclude that, since p, you're wrong. This would be to treat my judgment as infallible, which is unreasonable. Kelly's argument, though it is subtler than the direct argument, simply makes the same mistake in a subtler fashion. I believe that p, you claim the contrary. I treat the issue as already settled, and I don't allow your testimony to re-open the question, because I don't allow your testimony to challenge the reliability of my own faculties and abilities. This is to treat my abilities as infallible, which is just as unreasonable.

### **Epistemic Conclusions**

I now take myself to have finished the argument against moral self-sufficiency on strictly epistemic considerations. I've argued that we may reasonably trust our moral intuitions, but that that trust is only reasonable if its defeasible. As a further consequence, we must take our peers to have, in principle, as much access to the moral realm as we do. We are but one rational being among many, all of whom have access to the realm of moral facts, and some of whom have a similarly good *epistemic* claim on reliable access.

In this chapter, I've argued that we cannot dismiss the epistemic claim of others based merely on disagreement. The best argument for this dismissal - Thomas Kelly's - depends on suppressing the possibility that disagreement can give me evidence that my own mental abilities are malfunctioning or mistaken. This may work for conclusions for

which we have non-defeasible proofs, but for any domain where our access is dependent on trust in an entitled, defeasible faculty, we cannot so dismiss. Such dismissal in an entitled realm will depend on a supposition of *moral infallibility* - that I can resolve a contest between the reliability of my mental abilities, and the reliability of another's, through the unreflective application of those mental abilities. A reflective move, on the other hand, might clear things up. If I developed a good account of what made a person's moral judgments reliable, I could use it to sort between my judgments and others'. But I cannot settle the matter by simply by re-emphasizing my original judgments and the conclusions I drew from them, for it is the reliability of my process of moral reasoning which has put to question. Since each of us is merely one among many rational agents, many of whom have equally good epistemic claims on the moral realm, and since our claim and self-trust is defeasible, we must take disagreement as giving us some evidence to distrust ourselves.

This, I take it, is enough to quell the most significant standing objection on epistemic grounds. But now there is room for another sort of protest. "Sure, there are epistemic reasons to be morally humble," my opponent might say, "But epistemic reasons aren't the only reasons. There are moral reasons to stay true to one's intuitions. They are personal, they are part of your identity. To give up on them would be to give up on your autonomy." I now move to take up the particularly moral resistance to my conclusions.

## Chapter 4

### Autonomy, Understanding, and Disagreement

So far, I've focused on discussing the impact of epistemology on moral reasoning. I've argued that moral reasoning cannot occur in an epistemic vacuum; that considerations of self-doubt, attained through reflection on disagreement, should impact moral reasoning. Moral reasoning, as with all other reasoning, is dependent on the use of fallible human faculties and methodologies. What I've argued thus far is that there is no *epistemic* basis for moral self-sufficiency.

But perhaps the real basis for moral self-sufficiency lies beyond straightforwardly epistemic considerations. Perhaps there are peculiarly moral considerations that allow moral reasoning to escape from the usual epistemic constraints. Even if moral testimony and moral disagreement provide good *evidence* as to the truth, perhaps it is evidence that we ought ignore on moral grounds. Perhaps it is like evidence gathered through illegal police coercion: perfectly good evidence epistemically, but nevertheless wrong to use.

There seems to be no problem with delegating beliefs in the scientific sphere, but a different intuition seems to rule over moral matters. There seems to be a problem - a distinctively moral problem - with the soldier who shrugs off all responsibility for his actions, claiming that he was simply following orders. A person that simply follows everything their parent or superior says is at best still a child, and at worst a slave. There seems to be something within moral life itself that calls to each of us to reflect, consider, and decide for ourselves.

In this chapter, I will take up several distinctively moral arguments for moral self-sufficiency. First, I will consider the commonly held intuition that the use of moral testimony runs against personal *integrity*. Second, I will consider arguments from Robert Paul Wolff that any use of moral testimony is a violation of the constraints of *autonomy*. Third, I will consider the arguments of Philip Nickel, who claims that the overuse of moral testimony violates a special requirement for *understanding* in the moral domain.

I hope to show that, though each of these worries captures an important consideration for moral life, none of them is so powerful as to block the effect of moral disagreement. We are driven to weigh moral disagreement from our underlying commitments of any truth-oriented cognitive activity. Insofar as we are concerned with getting our moral beliefs right, non-epistemic considerations - integrity, autonomy, and understanding - may exert pressure on basic epistemic methods, but cannot obliterate them entirely. As long as we take ourselves to be cognitively fallible beings, and as long as part of our moral enterprise is to *get it right*, then we must be open to possible errors in our thinking and thus to the evidence of disconfirmation.

The claim of absolute moral self-sufficiency is an exaggeration of a more plausible consideration: that there is something wrong with giving ourselves entirely over to the command of another. Though I grant that there is something wrong with complete moral obedience, I will claim that the primary use of disagreement is not a form of obedience. It is, instead, a distinct process, part of an epistemic procedure of self-checking and corroborating our own mental abilities. Employing this procedure does not count as obedience; it is, in fact, virtually the antithesis. The drive to self-checking comes

from the very same values that lead us to abhor moral obedience: a drive to understanding, to accuracy, to self-perfection, and to greater responsibility for our moral beliefs. Moral obedience is a form of disengagement from the moral process. Using social sources of information to corroborate and dis corroborate our cognitive abilities is actually a form of increased engagement with moral reasoning, and a part of a mature rational approach to moral judgment. In fact, I will argue, the very values that lie under the importance of autonomy and understanding will drive us to actively seek out disagreement.

There is certainly something genuinely problematic with complete moral deference, with laying aside all moral thinking and proceeding entirely on authority and trust for others. Such deference violates one's responsibility to be a careful, informed, thoughtful moral agent. But using other agents as a way to check the reliability of one's own cognitive capacities is not a way of giving up that responsibility; it is in fact a way of fully living up to the responsibility of being careful, reflective moral agents. When we corroborate and dis corroborate our abilities, we are acting *more* responsibly, and more thoughtfully. The fact that the best practices for self-assessment involve corroborating our judgments with others doesn't show that corroboration violates autonomy; it shows, rather, that other people can help us become more perfectly rational, and more perfectly autonomous. Moral humility, I claim, is the part of the most rational and autonomous response to *all* the reasons and evidence at hand.

## Part I. Disagreement and Integrity

The first worry I'd like to take up is commonly felt, but rarely articulated in academic ethics: that, for reasons of personal integrity, we ought not weight disagreement. This is the least serious worry of the three, but a brief discussion of it will clear some useful ground.

Take Alan Donagan's case of the Austrian farmer. The farmer hears that WWII is coming, and he's called to arms by his government to fight on the side of the Axis. He's bothered by this; it seems to him that the Axis is in the wrong. He goes to his parish priest, and asks what he ought to do. The priest passes up the question, where it eventually reaches the Vatican in Rome, and the answer gets passed down: the Axis is on the side of right, and the farmer ought to go to war. But the farmer disobeys; he refuses because by his reasoning and his principles, the war is unjust. We might read this case in the following way: in this case, he has *very good reason to think the church is an epistemically better position than him*. Not only do they have better access to God, but they're educated folks who've thought longer about this stuff than him. He's told by his village priest that "neither they nor he, a relatively uneducated man, were in a position to make an informed judgment about the justice of the war; and that therefore the conclusion he had reached about it was doubtful."<sup>53</sup> But he sticks to his guns, and our intuitions say that he is to be admired for doing so, despite the fact that he is ignoring what he has every reason to think is very reliable testimony. Donagan's farmer is an archetypal hero of integrity.

---

<sup>53</sup> Donagan, Alan (1977), p 16

Why does integrity seem so important? First, we often take our felt beliefs and commitments to be partially constitutive of our *identity*. The second consideration we might call *sincerity* - that our moral beliefs are strongly connected to our emotions and perceptions, and acting against them violates some requirement for honesty. I grant that considerations of integrity are legitimate, but they cannot be trumping considerations. After all, when I see a bunch of KKK members on their way to lynch somebody from some particularly detested race, I don't think, "Excellent! What a victory for personal integrity against the prevailing attitude!" I think that the act is wrong, and I ought dissuade them or stop them in whatever way I can.

This example serves as a useful counterpoint to the Donagan case of the good-hearted farmer. Donagan's case is, I believe, subtly tilted, for *we* all know that in this case the farmer is on the side of right and the priest is in the wrong. We admire the farmer for maintaining the courage of his convictions, but his integrity is not the sole target of our admiration. We also admire him because he's right. If all we cared about was the integrity, then we ought equally approve and admire the racists for staying true to their felt beliefs, but we don't. It is more important that the racists not lynch than that they stay true to themselves.

Insofar as we take ourselves to be rational, and insofar as we take our beliefs to be aimed at objective moral fact, considerations of authenticity cannot be categorically trumping. What the overemphasis on moral integrity ignores is that we humans have a profoundly complex identity as both moral and rational agents. Insofar as we take ourselves to be rational, and to have rational moral beliefs, then we have at least two

commitments here. We have the considerations of moral integrity, which weigh on the side of staying true to our moral phenomenology, but we also have epistemic commitments to rationality, which weigh on revising our moral beliefs in the light of new considerations. This is why Donagan's farmer should not peacefully go to war, or peacefully go to jail; no matter how he acts, he should be emotionally *tortured*. He is under pressure from two warring commitments. Because our cognitive identities are sufficiently complex, we can be put in situations where we are at war with ourselves. Our moral beliefs are bivalent, and so are our moral identities. Loyalty to our complete nature as moral and truth-oriented agents requires we leave behind this one-dimensional notion of moral integrity.

## **Part II. Wolff on Autonomy**

Robert Paul Wolff offers a significantly more sophisticated path to moral-self sufficiency. Weighting disagreement in and of itself, says Wolff, violates the conditions of autonomy. A genuinely autonomous, responsible moral agent must reason for themselves and act from their own judgment. Relying on moral testimony in any way violates that autonomy. Wolff's argument is not only insightful, but also representative; it clearly articulates the worries that lie under much of the intuitive resistance to my claim of moral humility.

Wolff is surely right that some very particular uses of moral testimony are autonomy violations - outright, unthinking obedience. But obedience is not the only use



of moral testimony. I will argue that Wolff's argument cannot be extended to block appropriate uses of moral disagreement, and so cannot be used to fuel moral self-sufficiency. Furthermore, I will argue that the use of moral disagreement is appropriate *for the very same reasons* that unthinking obedience is inappropriate. When an agent thoughtlessly obeys, they are acting against the duty to be a thoughtful, responsible agent. When an agent uses agreement and disagreement to check their own moral reasoning - to corroborate and dis corroborate - they actually are becoming more thoughtful and being more responsible. They are extra-responsive to the reasons and considerations that impinge on moral action.

### **Wolff's account of autonomy**

Let me begin by defining some terms. I will attempt to conform my usage to that of the contemporary literature. I can be said to be "using testimony" an agent's testifying that p gives me some reason to believe p. "Moral deference," in contemporary usage, is a particular use of testimony.<sup>54</sup> Susan McGrath defines deference as a case in which "one holds a view solely because another person holds that view."<sup>55</sup> Problems surrounding moral testimony have also been examined as issues in *moral authority*. Elizabeth

---

<sup>54</sup> There is a slight quibble here: McGrath takes testimony and deference to be slightly different topics. For something to be testimony, it must be that another person actually speaks it. Her definition of deference does not require that another person actually speak their view; I can, for instance, come to think that my elder monk believes that ought not kill insects on the basis of his constant painstaking actions to avoid killing insects. I take it that the difference is not important to my present line of inquiry, since my interest has always been in the epistemic weight of the contrary judgments of one's peers, with little attention to how those judgments are communicated.

<sup>55</sup> McGrath (2011), p 113

Anscombe defines "moral authority" as "taking that somebody said something as over and above what you decide for yourself."<sup>56</sup> I will assume that the problem of moral deference and the problem of moral authority are interchangeable for the present purposes.

Wolff's discussion invokes a decidedly Kantian conception of autonomy to explain the problem with moral deference. An autonomous agent is self-governing and acts only on rules he gives himself, says Wolff. Certain uses of moral testimony count as acting on rules given by another, and so are violations of an agent's autonomy. Accepting the *command* of another, for example, constitutes a violation of autonomy, says Wolff:

Since the responsible man arrives at moral decisions which he expresses to himself in the form of imperatives, we may say that he gives laws to himself, or is self-legislating. In short, he is *autonomous*. As Kant argued, moral autonomy is a combination of freedom and responsibility; it is a submission to laws which one has made for oneself. The autonomous man, insofar as he is autonomous, is not subject to the will of another. He may do what another tells him, but not *because* he has been told to do it... Inasmuch as moral autonomy is simply the condition of taking full responsibility for one's actions, it follows that men can forfeit their autonomy at will. That is to say, a man can decide to obey the commands of another without making any attempt to determine for himself whether what is commanded is good or wise.<sup>57</sup>

We may not act on the command of another because we have been commanded; rather, we must determine for ourselves whether our action is good. Let us call this first requirement the requirement of *self-legislation*. The paradigmatic example of the failure of self-legislation is the unthinking soldier - the soldier who does anything his superior officer orders, solely on the officer's authority.

Wolff provides a second requirement, directed at the use of moral testimony:

---

<sup>56</sup> Anscombe (1981), p 44

<sup>57</sup> Wolff (1970), pp 13-14

The responsible man is not capricious or anarchic, for he does acknowledge himself bound by moral constraints. But he insists that he alone is the judge of those constraints. He may listen to the advice of others, but he makes it his own by determining for himself whether it is good advice. He may learn from others about his moral obligations, but only in the sense that a mathematician learns from other mathematicians - namely by hearing from them arguments whose validity he recognizes even though he did not think of them himself. He does not learn in the sense that one learns from an explorer, by accepting as true his accounts of things one cannot see for oneself.<sup>58</sup>

Let us call this second requirement the requirement of *moral understanding*. This is the requirement that when we self-legislate, we do so from reasons we see and by reasoning we have performed ourselves, understand, and accept. These two requirements for self-legislation and moral understanding together I will call Wolff's requirement of autonomy.

Wolff's argument is often taken as an argument for moral self-sufficiency on from considerations of autonomy. Wolff seems to claim that the only use for moral testimony is for presenting arguments that an agent then decides for themselves - that the testimony cannot have any weight in and of itself.<sup>59</sup>

---

<sup>58</sup> Wolff (1970), pp 13

<sup>59</sup> It is not precisely clear what Wolff's views on my own position would be. Wolff clearly forbids moral deference in its fullest form - that is, he explicitly forbids acquiring a new belief entirely from testimony. What he would think about partially downgrading a belief based entirely on testimony is not precisely clear. But I think that the spirit of Wolff's arguments are in the direction of forbidding my uses of disagreement. When I downgrade a belief based on disagreement, I am making a change in my moral beliefs based strictly on testimony, without understanding the reasons for the contrary beliefs. Insofar as Wolff seems to be postulating a strict requirement for self-containment, even this change, insofar as it is motivated only by testimony without any of my own avowal, it seems likely that Wolff would be against my suggested uses for disagreement.

## **The compatibility of evidence and autonomy**

Wolff is certainly onto something very important. It seems crucial that we do not give up our duty to be an active, contemplative moral agent, and Wolff presents a compelling account of why it is that we have that duty. But any reasonable autonomy requirement will not forbid the use of disagreement; it will only forbid unthinking obedience. In order to show this, I need to clarify the distinctions between certain healthy and unhealthy relationships between autonomous agents and testimony.

First, autonomy considerations can only forbid very particular relationships to testimony. In Wolff's Kantian language, autonomy considerations forbid giving oneself a law or rule based solely on the testimony of another. Autonomy considerations cannot forbid taking up simple, normatively neutral information through testimony, for this would invalidate most normal social procedures for information-gathering - including, for example, asking another person for directions.

One might be forgiven for momentarily supposing that autonomy considerations might forbid relying on any sort of testimony, rather than simply forbidding acquiring rules for action through testimony. After all, some interactions with testimony can look, superficially, like a form of epistemic slavery. After all, what I am claiming is that when a peer has challenged one's belief, one *must* reduce one's confidence in that belief. Isn't this a crime against autonomy? It seems that I have no say in the matter. In that way my considered belief in the matter is not entirely the product of my own thinking; another's beliefs have intruded into mine.

But this worry depends on an oversimplified notion of autonomy and freedom. Certainly, in the situation I describe, I am bound to alter my belief in a certain way because of the testimony of another, but this constraint is not anything special; rational agents are always bound to believe according to the weight of the evidence. Good testimony can sometimes rationally determine one's belief, but this is true of any form of good evidence. Imagine, for instance, that I am very worried about whether my friend has been stealing from me. Imagine I have a very reliable truth serum; I inject it into my friend and then, after waiting for the serum to take hold, I ask my friend if he's stolen anything from me. When he says no, I have good evidence that he isn't a thief, and this ought to sway my belief. Or, for a more everyday example, suppose that I have good evidence that my roommate is trustworthy and has no motivation to lie to me. If my roommate steps in from outside and I ask her how the weather is, and she replies, "It's quite nice and sunny out," the weight of the evidence is now for its being nice out. A rational being in this position, already committed to the rationality and trustworthiness of the roommate, ought to believe so.

The fact that a rational being is bound to believe according to the evidence can't be a violation of autonomy. To think otherwise is to misunderstand the complex relationship between autonomy and rationality. To be rational is to believe in accordance with the reasons. *Rational beings, in some very important sense, are not entirely unconstrained.* Their rational nature binds them; they must, on pain of irrationality, believe in line with the weight of the evidence. Our intuitions about freedom comport with this; if a contemporary scientist believes that  $2 + 2 = 5$ , or that the Earth is flat, the

scientist isn't free - he's dumb. It cannot count as a loss of autonomy to be rationally bound by evidence. First, taking evidence doesn't subvert autonomy in an improperly causal manner as say, storybook brainwashing might be. Taking evidence is a part of the functioning of our rationality; when we change our minds because of new evidence we've received, this is rationality working as it should. Whatever autonomy is - and I cannot hope to even begin to fully analyze the concept here - it must be compatible with the conditions of rationality. A rational autonomous being is free to decide whether or not to eat very fatty ice cream, and to choose between the sensual satisfaction of the ice cream and the health benefits of abstaining. But a rational autonomous being is not free to believe, after having read the scientific data, that the very fatty ice cream has no calories. That would be a failure of rationality.

The mere fact that the word of another partially determines my belief - that it constrains my belief - cannot in itself constitute a violation of autonomy, as long as that determination follows the rational pathways of evidential assessment. The fact that, in some cases, our beliefs are determined by testimony is not a special case; it is simply another way that rational beings are determined by the weight of evidence. Taking evidence is not the same as taking orders.

## **Obedience versus guidance**

What exactly is the difference, then, between the healthy way in which agents can be determined by evidence, and the unhealthy way in which agents forsake their autonomy through excess obedience? What is the nature of the evidence that may permissibly pass from one person to another? I do not think I can yet delineate precisely what may pass through testimony, for this will depend on developing a sophisticated account of what counts as a moral fact, as a moral rule, and as information, and an account of why some of these entities cannot be passed through testimony. That task is beyond the scope of the current work. I think we can, however, make some inroads into accounting what's going on in the clearest cases of problematic obedience, and show some uses of testimony that are clearly unproblematic. This will be enough to establish at least one unproblematic use of moral disagreement.

Let's begin with the paradigmatic case of bad obedience, the unthinking soldier. I will assume that the unthinking soldier does represent a genuine problem for autonomy. But what exactly is wrong with bad obedience? Is it that the unthinking soldier has no control whatsoever? No, says Wolff; the unthinking soldier begins by giving himself a rule of obedience, by which he ceases thinking and considering in certain circumstances. Furthermore, the initial rule of obedience is itself not unthinking - the soldier can clearly give reasons for it. "This officer is my superior officer." "We need to obey quickly to function as an effective unit." He can provide an explanation of why he is, in general, obeying, and he can in any instance of obedience tell you why he's obeying - though the latter explanations will presumably suffer from some monotony. What's missing is

something about the soldier's attitude and connection to the content of the particular rules that arise as he executes his orders.

Establishing precisely the problem with the unthinking soldier is actually rather tricky, for there are superficially similar cases that aren't as problematic. Take, for instance, the case of following somebody in a car. Imagine that my friends and I are coming back from a camping trip, in which we are divided into two cars. We wish to drive into town and meet up at a restaurant. Our cell phones have run out of power, and we are in unfamiliar territory. The sensible plan is for one car to be the leader, and the other the follower, and for the driver of the first car simply to decide, and the driver of the second car simply to follow. The situation is similar in important ways to the case of the unthinking soldier: there is an overall reason for following - coordination and the desire to eat together. And once the decision is made, the lead driver makes the decisions, for reasons, and the following driver follows those decisions, without understanding their particular content. Surely this case is morally unproblematic, but what, precisely, is the difference between this case and the unthinking soldier case?

Next, imagine that I come home on my birthday and my girlfriend blindfolds me and tells me to follow her. I ask why, and she says, "It's a surprise." She leads me out of the house, and I follow blindly. This case, innocuous as it might seem, is strikingly close to paradigmatic cases of bad obedience: I seem to be giving up on an understanding of the particular decisions I'm following; I don't have an understanding of the precise goal or end; my understanding of the end is vague at best. In the car-following case, I at least know what we're doing and why; in the girlfriend-surprise case, I don't even know that. I



simply trust her good will, and go. But this case is also obviously morally unproblematic. What sort of lovers - what sort of people - would we be if we refused surprises like this for reasons of autonomy?

So we have three cases of following directions of some form - two unproblematic cases, and one case of bad obedience. What is the distinguishing factor? It cannot be that the rule requires external input before it can fully determine a course of action, for all the following cases involve external input. Bad obedience cannot be characterized by a lack of understanding in the initial decision to follow, since the soldier can be said to understand why he is to obey, just as well as the following driver understands why he is to follow. Bad obedience also cannot be distinguished by the failure to understand the particular reasons for the subsequent actions and decisions, for in the car-following case, the follower also doesn't understand why he's turning right or left in each particular instance.

Perhaps, while the car-follower doesn't understand the particular micro-rules and decisions being used for each turn, the follower does understand the particular end being pursued fully, and precisely what the relationship is between that end, the constraints of the situation, and the methodology of following. But this can't quite be the crux of the problem either, for the surprising girlfriend case is one precisely where I, the boyfriend, don't understand the end being pursued at all - I merely have a general trust that my girlfriend has my best interests at heart.

Might it be, then, that the car-following case has limited obedience, and the soldier's case has unlimited obedience? Presumably, when we follow a lead driver, we

won't follow them *anywhere* - we will follow them as they turn left and right and stop, but we wouldn't follow them over the side of a bridge, through a crowded shopping mall, or into the side of a bus full of screaming schoolchildren. When we follow a car to a restaurant, we do not issue a blank check to the leader. This strikes closer to the heart of the matter, but it cannot quite be precisely it, for the unthinking soldier does not have to issue a blank check either. It is safe to imagine that an unthinking soldier wouldn't do absolutely *anything* - most probably would not follow orders to shoot themselves in the head, or shoot their commanding officer from behind, or to burn down their grandparents' retirement home. The unthinking soldiers may be unthinking only in a particular domain - about their actions in a foreign country or towards enemy soldiers or a particular ethnic group - and yet still be problematically unthinking, as they follow, without thought, their commander's orders to burn down villages, torture children, and slaughter civilians. The mere presence of a veto - an outer bound to the unthinking soldier's obedience - is not enough to make his actions unproblematic.<sup>60</sup>

The problem seems to lie not in the existence or non-existence of an outer bound, but in the precise nature of each individual instance of obedience; it is precisely in the attempt to surrender responsibility for their actions. If I follow my friend's car, and he runs a red light, and I run the red light also, I cannot excuse myself from responsibility by claiming that I was following another car. I am responsible for that decision; that is what separates the car-following case from the unthinking soldier case. What seems to differentiate the problematic and unproblematic cases is that in the car-following case I

---

<sup>60</sup> I am heavily indebted for this point to material from private conversation with Brian Hutler.

am being *guided*, and in the unthinking soldier case the soldier is *obeying*. In guidance, the leader is simply *suggesting* a course of action, which the follower assents or doesn't assent to in each case. The assent may be invisible, because it is so quick and nearly automatic; in car-following, the actions - turning left, changing lanes - are all so utterly routine that they are obviously morally unproblematic. But we can see that there is assent in each case, simply by imagining what happens when the leader car does something a little dicey - for example, running a red light. The following car has to decide whether or not they will follow, and the driver is responsible for that decision. In the problematic cases of obedience, I either act on a rule solely because it is endorsed by another, or endorse a rule on the sole grounds that another endorsed it. In the unproblematic cases of guidance I uptake information, and provide or withhold any endorsement myself. If my commanding officer issues an order for me to proceed to the north, obedience would be to follow with no further deliberation - endorsing it solely because my commanding officer endorsed it. Guidance would be to glean information from my officer's statement - for example, the information that my commanding officer thinks that it would be a good idea to go north - and use that information in a judgment of my own.

We can also see the distinction between obedience and guidance by looking at a paradigmatic case of guidance: tour-guiding. When I am being taken around town by a tour guide, and told to walk into certain places and eat certain things, what's being passed is information - the tour guide thinks this is a good place to see or a worthwhile place to eat. What I'm taking up is the fact about the tour guide's belief, and not entering the tour guide's judgment in as my own. When the tour guide says that the food at this restaurant

is delicious, I cannot in good conscience skip my meal, return home, and tell everybody that the food at that restaurant is delicious. What I can do is say that the tour guide thought that the restaurant was delicious, so it probably is; or, I can enter the restaurant for a good reason - that the tour guide thinks it's delicious - and then decide for myself. We can see the distinction between passing endorsement and passing information by looking at when we can and cannot defer responsibility to our tour guide. If the tour guide suggests that we might indulge in the local practice of cat-torturing, we are responsible for our actions if we do. If the tour guide lies to us and tells us that we may participate in this local practice of hitting a bag with a dead chicken inside, for meat preparation, and it later turns out that inside the bag was a drugged cat, we may justly pass on responsibility to the tour guide, because he has passed us false information, when we reasonably expected him to pass us good information.

The distinction between obedience and guidance captures something very important about where genuine autonomy problems lie, and where they don't. In the unproblematic cases, I am merely being guided - I am accepting only information from the leader, and then deciding for myself. In the problematic cases, I am obeying - I am accepting rules from the leader, which include information, decisions about appropriate ends to pursue, and decisions about the best means to achieve that ends. I am, in some sense, letting them decide for me. Guidance may resemble obedience closely, but they are crucially distinct. I may modify my beliefs and actions in response to testimony, but this in and of itself does not show problematic obedience. What matters is whether my own judgment and endorsement is interposed between the testimony and the action. When

another provides only information, and I provide the rules that connect that information to actions, then any following I'm doing is unproblematic, though it may superficially resemble obedience.

Armed with the obedience/guidance distinction, it now seems easy to explain the blindfold-surprise case. There, we are placed in a special and very vulnerable relationship to our leader - we make them, for a period of time, the source of a tremendous amount of information that we would normally gather ourselves. And when they pass us bad information, we can pass on the responsibility to them. But we cannot pass on responsibility when they suggest bad rules.

Guidance may resemble obedience closely, but they are crucially distinct. I may heavily modify my beliefs and actions in response to testimony, but this in and of itself does not show problematic obedience. What matters is whether my own judgment and endorsement is interposed between the testimony and the action. When another provides only information, and I provide all the rules that connect that information to actions, then any following I'm doing is unproblematic, though it may superficially resemble obedience. Suppose I am on an airplane and a man in front of me begins to choke; I get a doctor on the phone and he walks me through a tracheotomy. It may appear that I'm obeying; after all, the doctor says, "Now cut a one-inch hole underneath his adam's apple," and I do it immediately, but here I am only being guided. I have already decided of my own cognizance to save this man's life by the most expedient means necessary; the doctor merely provides information about the means. Cases of obedience and cases of guidance may be virtually indistinguishable in the moment of action, but we can tell the

difference by the attitude of responsibility the agents have towards their ends and rules. In guidance cases, I may pass responsibility about particular pieces of information, but I cannot pass responsibility about my decision to act from that information. Obedience cases are marked precisely by their attempt to abstain from accepting responsibility for the decision to act.

### **The indirect pathway to doubt**

There is surely something wrong with the sort of outright moral deference that marks unthinking obedience. I do not take myself to have fully explicated what that wrong is, and where its boundaries are. Perhaps there are some cases where some degree of obedience is permitted, some forms of uptaking endorsement or moral rules that are permissible; the matter seems terribly complex. What I take myself to have already shown, however, is that there is a category of obviously permissible interactions with testimony: uptaking non-moral information. Furthermore, there are many situations which superficially resemble cases of bad obedience, but in which we actually only uptake non-moral information - cases of guidance. What I intend to show now is that there is at least one use of moral disagreement where we acquire through social pathways strictly non-moral information – where we use the moral testimony of others as guidance, not obedience. This is the use of moral testimony for corroborating and dis corroborating our cognitive abilities - for self-checking.

Disagreement can give me a reason to suspect the reliability of my own abilities. Evidence that my faculties are potentially unreliable is not a rule for action; it not a moral fact of any sort.<sup>61</sup> It is information about a non-moral matter, namely, the proper functioning of my cognitive faculties. Thus, even if we subscribe to the strongest prohibition against uptaking any form of endorsement or moral rule, there is still a pathway for moral disagreement to permissibly affect our confidence.

First, let me distinguish between two basic uses of testimony. I can use testimony as a reason to believe what the testifier believes; I can also use testimony to establish what a testifier believes. The former use may be forbidden, but the latter is clearly permissible. It cannot be problematic to uptake and use the information that my commander *believes* that we ought to burn down the village of innocents, even as I plot to sabotage his plans. There are, accordingly, two ways in which testimony can be used to modify my own belief; let me call them the direct pathway, and the indirect pathway. In the direct pathway, I come to alter my belief immediately in accordance with received testimony. In the indirect pathway, I use the testimony to establish what my interlocutor believes, then use the fact that we agree or disagree as potential evidence of our respective reliability or unreliability.

We don't often separate these two uses of testimony in everyday empirical matters, since both uses of empirical testimony are permissible, but the distinction is crucial for moral testimony. The distinction between the two paths may seem

---

<sup>61</sup> I am aware that I am relying heavily on a distinction between information and endorsed rules, without having provided an explicit account of the distinction. I do not possess such an account, and I do think there are many unclear boundary cases, like facts about aspects of character like dignity or insensitivity. But the process-reliability facts seem clearly informational, and clearly outside the space of norms, endorsed rules, and any of their kin.

superficially fine, but I will claim that they are strikingly different. In the direct path, I am *obeying*, and in the indirect path, I am merely being *guided*. In the indirect path, I am uptaking an endorsement or rule of action directly. In the indirect path, I am using testimony to glean a piece of morally neutral information: that an interlocutor I take to be reliable on the subject disagrees with me. The distinction between the two pathways is everyday. When I'm doing my math homework, I am working under the requirement that any answers I commit myself to are my own. It does violate that requirement to simply copy my partner's work and conclusion. But it surely doesn't violate that principle to check my work against other students, and to doubt my work (and theirs) when there is disagreement.

The difference between the two pathways rests upon a difference between two processes of reasoning. First, there is the process of moral reasoning, which goes something like: "It certainly seems to me that suffering of any sort ought not be brought about, unless it is to prevent other, greater suffering. And when I eat veal, I am participating in a system that systematically creates suffering, but to an end that doesn't prevent other suffering. It only gives me a mild pleasure. So I must not eat veal." The moral reasoning process depends on the use of a number of entitled, defeasible substantive cognitive faculties.

But there is an independent reasoning process I might engage in to check on the reliability of those cognitive faculties used in the first process. I can note a disagreement between myself and another trustworthy moral peer on this topic, and so acquire evidence



that my reasoning process is possibly unreliable. Call this second reasoning process the self-checking process. I claim that:

- A. the self-checking process is distinct from the moral reasoning process.
- B. the self-checking process does not yield a rule of action or a moral claim of any sort, but a proposition of empirical fact.

The self-checking process depends only on using a morally neutral, empirical fact through testimony, and no reasonable autonomy requirement can block the use of morally neutral information. The self-checking process starts in a different place from the moral reasoning process: the moral reasoning process begins with a moral intuition, the self-checking process begins with an empirical observation about what somebody else said. The moral reasoning process yields a rule for action, where the self-checking process yields an empirical claim about the reliability of an ability or process.

The results of the self-checking process can modify my confidence in the results of the moral reasoning process, but this does not mean that they are the same process, or that constraints on the moral reasoning process must also apply to the self-checking process. My all-things-considered final judgment of what to do depends on the results of several processes - both the distinctively moral first-order reasoning process, and the non-moral second-order self-checking process.

Using the direct pathway for moral testimony would count as obedience, because I am endorsing a rule solely on the grounds that another endorses it. The direct pathway

is an attempt to skip over the moral reasoning process. But using the indirect pathway is merely a form of guidance. I am using the fact that another endorses a particular rule as evidence for my own reasoning. I am not skipping over of a moral reasoning process, but performing a complex, engaged process of self-checking. These are two distinct processes, one essentially moral and directed towards receiving rules of action, and the other essentially non-moral and directed towards generating information about the reliability one's own mental abilities and faculties. Autonomy considerations can forbid an attempt to replace the moral reasoning process with deferral to testimony, but autonomy considerations have no application to the morally neutral self-checking process.

Using disagreement as part of the self-checking process will not run afoul of Wolff's autonomy requirement for moral reasoning, because self-checking isn't moral reasoning. The self-checking process will, however, impact my confidence in the results of the moral reasoning. This may seem paradoxical, but it is not. This conclusion depends on the claim that there are empirical facts that lie under and buttress my confidence in my moral reasoning; thus, non-moral reasoning can infect my confidence in my moral reasoning. But surely this claim is correct. Surely empirical evidence can impugn my trust in my own judgments - from the discovery that I've been on drugs, to the discovery that I have a degenerative mental disease, to the simple discovery that even though I thought I was good at calculus and did those calculations very well, I failed three calculus tests in a row. Surely, if I were to find out that I had been under the influence of drugs, or post-traumatic stress syndrome, when I passed moral judgment on my friends, I would

have reason to worry out and doubt those judgments. Since in all reasoning, including moral reasoning, we are putting our trust in defeasible reasoning processes and abilities, it seems difficult to see why there *couldn't* be empirical claims that could impugn that trust.

### **The independence of the direct pathway**

One might then suspect that the indirect pathway, though it is distinct from the direct pathway, depends on a hidden use of the direct pathway. After all, don't I have to use the fact that my interlocutor is morally *reliable* to get to the belief that their disagreement means something? And isn't leaning on the moral reliability of an interlocutor simply granting them direct access to my moral beliefs?

I grant that the indirect pathway relies on invoking a belief in the interlocutor's reliability, but this invocation does not constitute problematic obedience. It is not the same as directly uptaking a rule. We can see this in two ways. First, a key marker for obedience cases is the agent's attempts to pass responsibility for his action on to his testimonial source. What seems to mark the unthinking soldier cases is that the unthinking soldier abandons his responsibility for having correct beliefs. There is no attempt to pass responsibility in any form in self-checking through testimony. In fact, when we use interlocutors to probe ourselves for possible errors, we are taking up even more responsibility in the pursuit of correctness.

The independence of the pathways is clear when we look at the rule of action that's actually doing the normative heavy lifting. In the obedience case, action follows from a rule I acquired through testimony. But in using disagreement for self-checking, the action I take - modifying my belief, refraining from action - follows from a rule of action I give myself, which takes as inputs empirical facts about disagreement.

Suppose that I am considering eating this piece of chicken, which I know was raised on a commercial farm. I think this is fine, because I am operating under the following rule:

Omnivore Rule: Gustatory pleasure can justify some animal suffering, because animal suffering is insignificant (depending on the type of animal).

Somebody else claims that this is incorrect, and that the proper rule of action is:

Vegetarian Rule: No amount of gustatory pleasure can justify animal suffering, so you ought not eat animals, especially those that have suffered.

If I were to begin to believe the Vegetarian Rule immediately, based solely on testimony, then this would be taking the direct pathway; it would be a violation of Wolffian autonomy.

But I can take the indirect pathway instead. I note that there is disagreement. I find the omnivore rule intuitive, and my interlocutor finds the vegetarian rule intuitive.

There is a disagreement, and since both interlocutors are trustworthy, and since I have no ready explanation for the disagreement, I have some evidence to suspect both interlocutors, including myself, of unreliability. Now suppose I act using this evidence using the following rule:

Faculty Trust Rule: When there is evidence that a cognitive faculty is or might be unreliable to some degree of probability, I ought to reduce my trust in the judgments that relied on that faculty by an appropriate measure.

The indirect pathway relies on the Faculty Trust Rule. And it is true that if I endorsed the Faculty Trust Rule itself merely on the basis of testimony, I would be sacrificing my autonomy. But the Faculty Trust Rule doesn't have to come to me through testimony; it is a rule I can give myself. In the autonomy-preserving case, I act from Faculty Trust Rule, which I give and endorse myself, and which takes as an input the information that another person believes the Vegetarian Rule. In this case, the change in action comes from an application of my Faculty Trust Rule, not from a direct transmission of the interlocutor's Vegetarian Rule into my belief system. In the indirect pathway, I am not adopting the vegetarian rule based simply on testimony, without endorsing or understanding it. I am *using* the fact that another reliable person thinks the vegetarian rule as a data point, and engaging with that data point using a rule that I endorse and understand.

Finally, we might worry that the Faculty Trust Rule's plausibility depends on assigning the Vegetarian Rule some evidential weight from testimony. After all, the

process of corroboration and discredoration only makes sense if we treat an interlocutor's believing something as in some way impacting on its likelihood to be. And it's true - but what we're granting here is simply the thesis that moral matters are cognitive, and that we take others to be reliable sources on cognitive matters. The purpose of the foregoing discussion of obedience was to show that the problem of unthinking obedience was very particular; it depends on *endorsing* a rule on the sole grounds that another endorsed it. Here, in no way am I endorsing the Vegetarian Rule merely on another's endorsement; I am merely using the fact of another's having endorsed it as a data point for my own reasoning process.

This may seem like hair-splitting. It is not. I am trying to show that the process of obedience and the process of discredoration entirely different; that what we find abhorrent about blind obedience is nowhere to be found in discredoration. When we obey unthinkingly, we take ourselves out of the reasoning process. We cannot give any explanations for the rules we follow; we are not involved in the process of weighing and deliberating and deciding. We make at most a single decision - to obey an authority - and afterwards step out of the picture. We do not make ourselves responsible for the endorsement of a rule, because we played, at most, a paper-thin role in reaching it. The process of corroboration and discredoration using social information is the opposite. When we check our reasoning against others, we are buried in the reasoning process. We are fully engaged in two distinct forms of reasoning - first, forming our mind independently about the moral matter, and then checking the reliability of that first process through another process. Both processes proceed in the full light of

understanding: I know why I've decided that eating meat is allowed (because I think animal lives are, in the end, is genuinely less important than human desires), and I know why I hesitate to act on that belief (because not everybody agrees with me, and because this is a reason to worry). I am in full contact with the reasons and rules involved; I am fully a *participant*, and not a resigned bystander, in the reasoning process. The unthinking soldier becomes a mere conduit for his officer; the officer's endorsement bypasses the soldier's reasoning process and leads directly to action; his rules are merely his officer's rules with almost nothing added from himself. When we worry, listen to disagreement, and doubt, our final, modified, suspicion-laden rule of action is the product of two processes of reasoning, in both of which we have fully participated.

### **Diagnosis: bivalence and alienation**

If the process of corroboration and discorroboration is not essentially a moral matter, why does it trip alarms in the vicinity of autonomy? Let me offer a diagnosis. The autonomy-based worries about disagreement depend on drawing a compelling, but false, dichotomy between our moral commitments and our epistemic commitments. The autonomy worries arise from seeing the sort of rational, epistemic pressure from discorroboration as an alien and invasive presence in our moral lives. The view depends on seeing our relationship with our moral beliefs as more genuine and integral than considerations of cognitive reliability. When procedures like self-checking and discorroboration get in between my private moral experience and my all-things-

considered beliefs, it may seem like I'm alienating myself from my moral commitments. But this view suppresses the complexly cognitive nature of moral beliefs, and our complex nature as rational autonomous beings. Our moral commitments are an essential part of our personal identity, but so are our epistemic commitments. We are cognitive beings as well as beings soaked in moral feeling. When we note disagreement and lose confidence in ourselves, we are not being alienated through an invasion from without; we are in conflict with another part of ourselves.

The autonomy worry arises from a one-dimensional view of moral beliefs: the view that our moral beliefs are simply products of our freedom, as expressive of our nature and phenomenal experience of the moral world. As long as we are moral cognitivists, we must view moral beliefs as essentially *bivalent*. We are committed both to a moral belief's being expressive of ourselves *and* as getting it right of facts independent of ourselves. The first set of commitments drive us to keep our moral beliefs private, to seal them off from social input. But the commitment to correctness drives us in the opposite direction; it drives us to procedures of corroboration and discorroboration, to general epistemic principles that obtain of any cognitive domain. Neither of these commitments are alien; both sets of commitments arise from the nature of moral beliefs themselves. Insofar as we take our moral beliefs to be aimed at truth, we commit ourselves to using procedures oriented towards accuracy and reliability.

It does seem that there is friction between the commitments to moral expressiveness, and the commitments to moral accuracy, but this friction is internal to our body of commitments. We have divided loyalties because we are divided beings. We are



rational and autonomous; we wish to rule ourselves by rules we give ourselves, and we wish to give ourselves rules that are in line with our felt, phenomenally vivid moral experiences; but we also wish to give ourselves the right rules. These commitments can run counter to each other. It is, in fact, inevitable, so long as our moral feelings are simultaneously personal, purportedly cognitive and fallible.

The same friction can be seen elsewhere. I wish, for example, to be with a romantic partner that I feel love for. But I also wish to love properly. I have learned - through trial, error, and lots of advice from friends - that I tend to be attracted to sadistic narcissists, and this attraction easily blooms into love. So I walk a difficult path - I try to stay true to my feelings, but also stand back from them, make sure they're right. I cannot inhabit my loves without conflict, for my loves are enmeshed in complex, sometimes contradictory commitments. This conflict is not an invasion from without, for it arises from the fact that there are pressures from different directions laid over my love. The friction is not a sign of alienation, but of the complexity and difficulty of human commitments.

### **First-order content and second-order reflection**

I am certainly not arguing that all uses of testimony are legitimate. I grant that some autonomy considerations are legitimate and that direct uses of testimony count as violations of autonomy. And by forbidding the direct path, we are forbidding many substantive uses of testimony. Only via the direct path could I acquire a new first-order

moral judgment, and do it while circumventing my reasoning process. The indirect path can only manipulate the second-order degree of confidence I have in a first-order moral judgment I've arrived at through my own reasoning. In forbidding the direct path, we forbid the wholesale acquisition of rules from other people; what's left is the manipulation of our self-confidence in the reliability of our moral reasoning. The latter manipulation cannot be forbidden by autonomy considerations precisely because it is not an essentially moral matter.

If we grant Wolff's view about the problem of obedience, and if my analysis holds, we get two very interesting results. First, the effect of testimony will frequently be asymmetric, especially for agents with a high initial degree of self-confidence in their moral reasoning. Since disagreement can only provide us with a second-order reliability assessment, agents with a very high initial self-confidence can only be moved towards lowered self-confidence through disagreement. But for agents with less than complete self-confidence, second-order reasoning can *restore* trust as well. Imagine, for example, that I am a young Dickensian orphan, who is very morally conscientious, and attempts to reason perpetually. The person who I trust, the figure of authority, constantly tells me that I am wrong, that I have reasoned in error, and that my beliefs are incorrect. I may come, quite reasonably, to distrust my reasoning process, and distrust my conclusions. Later in life I discover that most of my early judgments are also held by a vast number of very good, sensitive people, and that in the opinion of many, that early master of mine was a lying, deceptive, manipulative tyrant. So my confidence in my judgments might be restored.

More importantly, the impact of disagreement will be entirely second-order. We have a complex, double-stanced relationship towards our moral judgment. Granting Wolff's block against obedience, we cannot use testimony in the formation of new first-order moral beliefs. We are closed to all forms of moral testimony *for the purpose of generating positive moral content*. But we also have a reflective stance on our own cognitive abilities, including the abilities used in those private deliberations. And as rational, fallible, self-aware agents, we must admit that our cognitions sometimes go wrong. The role of testimony will only be in adjusting the degree of confidence I have towards my first-order beliefs.

This dual reflective stance shows up in other places; most of us hold such a position towards our judgments of romantic love. We (at least we modern, western folk) seem to hold a principle of the autonomy of love. We think that when we love somebody, that love must be arrived at through a private process of deliberation and feeling. Nobody can tell me *who* to love; that is something that only I can come to discover for myself. But, at the same time, my process of finding love, falling in love, and accepting my loves, is subject to input from other people. Though only I may come love another, others can certainly help me figure out when I've loved badly, or when I'm mistaken about my feelings. I can be reminded by others that my judgment is impaired, that I'm recently out of relationship and my emotions are wild, or that I've had too much to drink that night and am clearly drunk and so shouldn't get married. Only I can decide who I love, but others can suggest to me that my judgment is, at the moment, impaired.

## **Rational self-perfection**

What I've been arguing is that any reasonable theory of autonomy *must allow a rational autonomous being to uptake information without thereby violating her autonomy*. I've suggested a distinction: that autonomy violations will involve uptaking something in the vicinity of norms or rules or goals through testimony, whereas uptaking morally neutral information cannot count as an autonomy violation. When I use the existence of disagreement to reflect about my own reliability, I am a full participant in the reasoning process. I am acting on information I have gleaned with the help of others, by using the faculty trust rule which I understand and endorse myself. This pathway is autonomy-preserving.

I can put things a little more clearly if I help myself to more of the Kantian framework which seems to lie underneath Wolffian autonomy worries. Let me suppose, for the moment, that there is a very strong connection between rationality and autonomy. We are rational beings, and in reasoning well, we are autonomous. The connection is likely the animating force behind Wolff's worries. The unthinking soldier is not reasoning at all; he is submitting. He is giving up on his duties as a rational being to think, consider, and take evidence and reasons into consideration. But the epistemically motivated doubter is not giving up on thinking. She is reasoning more and insofar as she is reasoning from good epistemic principles, she is reasoning *better*. We can see this by examining the ends that move our two agents to action. The unthinking soldier is moved by end that comes directly from his commanding officer. It is the officer's end, and the

officer's justification via that end, that moves the unthinking soldier. The doubter, on the other hand, is moved by an end that does not come through testimony. The doubter's proximate end is checking the reliability of her own cognitive abilities (which likely derives from the doubter's ultimate end of having correct beliefs). The doubter does not acquire these ends from the testifier, and we can demonstrate this simply: a doubter can use the testimony of an interlocutor who does not share the end of checking their cognitive abilities.

Once we begin to take into account the moral beliefs of others - not by submitting to the command of others, but by becoming aware of the reasoning of others and using it as evidence - we become more rational. We are engaged in further rational activities: we are triangulating, we are corroborating, we are error-checking, we are debugging. We are engaged in an activity of increasing the reliability of our judgments. We are searching for evidence of our errors, evidence which may lead us to feel out where we should think again, where we should worry, where we should theorize. What this shows is that not all instances of interaction with testimony are bad. Certainly, there is such a thing as autonomy-destroying obedience. But there is also the proper use of other people as rational sources. And if in so doing we are becoming more accurate, more well-informed, more self-conscious reasoners, then we are increasing our autonomy. It is true that we may allow ourselves to blindly obey other people and so lose our autonomy, but it is also true that other people can help us to become more autonomous, as long as we use evidence of their testimony thoughtfully and properly, and not lazily. The unthinking

soldier has abandoned interaction with standards of correctness, while she who attends to disagreement is thereby pursuing correctness more fervently

### **Part III. Moral Understanding and the Right Sort of Reasons**

Wolff's account focused on explaining the wrong of moral deference with references to very Kantian notions of autonomy. Considerations of autonomy were supposed to ground requirements for agents to reason, understand, and choose independently. The other significant branch of contemporary discussion of moral deference does without invocations of Kantian autonomy; these other accounts explain the wrong of moral deference by invoking a duty for moral understanding. I will now turn to these accounts of moral understanding.

There is, according to these accounts, a duty to understand one's own moral beliefs and judgments. Moral deference and other excessive uses of moral testimony are wrong because they replace moral understanding with trust. When an agent undertakes actions solely on the say-so of another, they may get the action crudely right, but they do not do so with a full understanding of why the action is right.

I think there likely is an independent duty for moral understanding, and I think that some of the contemporary accounts of the duty are as plausible as the autonomy accounts.<sup>62</sup> But I do not think that any reasonable version of that duty will threaten my

---

<sup>62</sup> Either a Wolffian autonomy account or a moral understanding account is capable of independently explain the wrong of moral deference. I suspect that the truth is that both accounts are correct, and that

arguments. In fact I think that the most plausible account of the duty to understand will actually show that we are *required* to attend to disagreement by the *very same values* that animate our duty to understand. The duty for moral understanding emanates from a duty for moral self-perfection, for transforming ourselves into the most accurate, reliable, morally sensitive beings we can. Using others to check and reflect on our own reliability is another vital part of that process of moral self-perfection.

### **Hopkins and grasping the moral grounds**

Let us presume that there is something wrong with outright moral deference and that what's wrong is, at least in part, that outright moral deference engenders a systematic lack of moral understanding. But does weighting disagreement also engender a lack of moral understanding? After all, when I attend to disagreement with a peer, I am invoking a substantive degree of trust in that peer. I am allowing the fact that my peer believes *p* to influence my degree of belief that *p*, even if I do not understand their reason for believing *p*. Does this count as a form of problematic deference? To answer, we will have to take a closer look at why there is a special duty for moral understanding.

Robert Hopkins offers an excellent overview of the problem of moral deference in his paper, "What is Wrong With Moral Testimony". He raises most of the arguments given for the problem of moral testimony and disposes of most of them quite nicely. He argues that, if there is something wrong with moral testimony, it must be because

---

agents are subject both to requirements for autonomy and independent requirements to understand founded on epistemic norms.

acquiring a belief exclusively via moral testimony grants no moral understanding. When you settle the matter on testimonial grounds, writes Hopkins, "you have reasons for your belief but not moral reasons for it."<sup>63</sup> Hopkins suggests the following requirement on moral beliefs:

The Requirement: having the right to a moral belief requires one to grasp the moral grounds for it.<sup>64</sup>

Does Hopkins' Requirement present a problem for the use of disagreement for disconfirmation?

Hopkins' Requirement, if it is true, blocks only the most outright form of moral deference, and does not block the use of moral disagreement for disconfirmation. This is because, as I've argued above, *the negative use of moral disagreement is a form of moral deference*. When I acquire self-doubt through the observation of disagreement with a vegetarian, I do not come to think that it is moral or immoral of me to trust myself, or to continue eating meat. The content of the beliefs I acquire through the disconfirmation process are strictly about myself and the reliability of my own faculties. I acquire evidence that my moral faculties might be malfunctioning, and I withdraw some degree of trust in that part of my judgment on epistemic, and not moral grounds. Any new belief I have acquired are strictly epistemic beliefs, and don't fall under the scope of Hopkins' Requirement. Furthermore, *I fully understand the reasons for reducing my degree of confidence*. I understand the relationship between disagreement, disconfirmation, and

---

<sup>63</sup> Hopkins (2007), p 19

<sup>64</sup> Hopkins (2007), p 20



self mistrust. The use of disagreement for discreditation escapes Hopkins' understanding requirement in precisely the same way that it escapes Wolff's autonomy requirement. The legitimate, non-deferential use of moral testimony is via the indirect route: as evidence of potential malfunctioning. Thus, though I grant Hopkins that his Requirement is moderately plausible, and would block moral deference, I do not grant that it blocks all uses of moral testimony. It does not block the non-moral self-checking process.

### **Nickel and the Requirement of Recognition**

Hopkins' requirement is very strong, and demands that we possess understanding for each and every moral belief we have. Philip Nickel presents what I take to be a more sophisticated view. At the center of the view is a more moderated version of the understanding requirement, in which an agent has *in general* a duty to understand their moral beliefs, though his duty admits of occasional exceptions. Nickel's account strikes me as the most plausible version of a pure understanding requirement, and I think if we accept it, we will find even stronger support for the use of disagreement. I hope to show that Nickel's account is not only compatible with my own, but that Nickel's account *requires* that an agent also attend to disagreement and agreement to assess their own reliability. The self-checking process I describe and the moral understanding Nickel favors are both means to the same end: ways for an agent to strive for moral self-perfection.

Nickel's view is that:

Morality aims at guiding action rationally, i.e. *from* a recognition of the relevant moral requirements. A moral agent must be responsive to morality as such...<sup>65</sup>

This is a requirement of morality, claims Nickel; part of what it is to act morally is to act from an understanding that the act is moral. The use of moral testimony typically provides correct moral belief without understanding, says Nickel. In these cases, "although moral testimony may give rise to a correct moral belief in these cases and thus 'work out' in the crudest sense, it does not provide a basis for morally good action".<sup>66</sup>

Nickel's argument is carried out first through a series of intuitive examples.

Nickel asks us to consider the following cases:

CASE A: My friend is addicted to heroin, and asks for money. I tell him that I will lend him the money as soon as he gets his life together, and my justification is that my mother told me to say this.<sup>67</sup>

About Case A, Nickel says surely the friend should be unsatisfied, even if I defend my claim with explanations of my mother's greater age and experience. There is something, says Nickel, lacking about my justification if I offer no independent support for my moral beliefs, but only defer to my mother's judgment.

---

<sup>65</sup> Nickel (2001), p 256

<sup>66</sup> Nickel (2001), p 260

<sup>67</sup> Nickel (2001), p 256

CASE B: A child is trained to act in accordance with certain rules, not by explanations, but by behavioral conditioning and punishment. Once the child is grown, she is unmoved by moral concepts. But she continues to perform moral-seeming actions (like returning borrowed objects) because she feels the force of her early conditioning.<sup>68</sup>

About Case B, Nickel says that we have the intuition that something is missing from this action, which is that the action fails to come on the basis of recognition. Thus Nickel presents us with a strongly worded Recognition Requirement:

It must be the case that morality requires one to *act* from an understanding of moral claims, and therefore to *have* an understanding of moral claims that are relevant to action.<sup>69</sup>

If our intuitions about Case A and B are true, says Nickel, then the Recognition Requirement must hold.

Nickel's version of the duty of moral understanding is significantly weaker than Hopkins', for very good reasons. Hopkins' thesis is that moral understanding is a requirement for moral action on a case-by-case basis; it admits of no exceptions. Nickel makes room for some special cases, in which a moral agent who in general seeks and possesses moral understanding, occasionally steps aside and defers to testimony. Nickel here is explicitly taking on board Karen Jones' work on moral testimony. Jones argues that moral testimony can sometimes be useful as a corrective for bias. A moral agent who

---

<sup>68</sup> Nickel (2001), p 257

<sup>69</sup> Nickel (2001), p 257

in general possesses moral understanding may sometimes be forced to acknowledge a blind spot in his own sensitivities. For example, Jones suggests that a morally sensitive man might still be insufficiently sensitive to some of the difficulties of a woman in the workplace, especially to the pervasiveness of certain pressures and their moral import. In a case where such sensitivity was vital, a morally sensitive man may choose to defer to the judgment of somebody more sensitive to that particular matter.<sup>70</sup> Somebody who uses testimony occasionally and judiciously, as a corrective to bias, is still acting "with a conception of the relevant reasons in mind," even if they're unable to apply the relevant reasons accurately in each case, says Nickel. "It is only when one has a more global inability to grasp relevant reasons, or when one refuses to do so, that one cannot act morally well."<sup>71</sup>

Nickel defends the Recognition Requirement as not only intuitive, but rational and reasonable. First, he claims that understanding a claim is vital to being able to apply the claim generally. When I depend on reliable moral testimony in one case, though I may get the action right in that single instance, I won't be able to reliably get similar cases right.<sup>72</sup> Second, says Nickel, if I don't understand the claims underneath my belief, my actions will be more error-prone. Successfully applying moral testimony requires a static situation. Typically, I have to communicate the situation to my advisor, receive their advice, then deploy the advice. Even if the advisor's testimony is correct, my acting correctly depends on the situation's remaining unchanged in its relevant details between

---

<sup>70</sup> Jones (1999)

<sup>71</sup> Nickel (2001), p 264

<sup>72</sup> Nickel (2001), p 261

receiving the testimony and acting. And, of course, lacking understanding, I don't know which parts of the situation are relevant. On the other hand, says Nickel, moral understanding lets me respond flexibly to changing situations. If I know what's important, I will be able to tell which changes to the situation are unimportant, and which changes to the situations are highly relevant. An agent "must try to ensure that she is in a proper position to act from a recognition of what morality requires, and she is culpable for failing to do so." This shows that there is a "substantial epistemological duty to be able to tell what counts as a justification for a variety of moral claims," says Nickel.<sup>73</sup> If an agent doesn't strive for understanding, she will fail to be flexible, and fail to retain her moral correctness in a fluid, ever-changing world. For Nickel, then, the connection between the duty for moral understanding and moral action is not constitutive, but pragmatic. Moral understanding is required for me to get it right more often.<sup>74 75</sup>

Nickel's view that the duty of moral understanding is based ultimately on the duty to be morally correct sits very well the Jones-style bias cases. The cases where Nickel allows a deviation from the strong case-by-case requirement for moral understanding are precisely those special cases where occasional deviations from the rule of moral

---

<sup>73</sup> Nickel (2001), p 261

<sup>74</sup> His account could even be used to show how acting independently would be better, even if in that particular case moral deference would lead to more morally correct action. This is because acting from one's own understanding is part of the process of developing one's understanding. Even if, in one's youth, one's parent is clearly more reliable than oneself, part of the process of becoming a morally able agent involves frequently striking out on one's own and working from one's own understanding. The errors one pays by avoiding deference in some cases will pay off in one's greater abilities down the line.

<sup>75</sup> In fact, it seems to me that Nickel has established something with these arguments slightly weaker than his initial claim. His language of his opening claim suggests that he will argue for a *constitutive* relationship between moral understanding and moral action, in general, where his arguments merely seem to establish that there is an *instrumental* relationship between moral understanding and moral action. My analysis of Nickel's position here relies on the content of his argument, rather than the language of those earlier passages.

understanding will lead to getting it right more often. An agent that never seeks moral understanding will be unable to get it right in most cases; moral deference, thinks Nickel, can never be as reliable as moral understanding. But an agent that possesses moral understanding in most cases can permit themselves a few deviations at the limits of their sensitivities, in those cases where doing so gets it right more often. Importantly, says Nickel, those deviations are performed with a conception of the relevant reasons in mind. Presumably, what Nickel means here is that if take myself to be less than perfectly sensitive to the moral issues of women in the workplace, I am still working with the right reasons in mind. I am acting in order to treat others justly and fairly. I simply lack the ability to figure out the best way to achieve those goals in this particular situation, because I lack some relevant sensitivity.

For Nickel, the wrongs, demands, and duties associated with moral understanding gain their normative force as part of the means to getting it right. The duty here is a practical one: we do not have a duty to achieve understanding for its own sake, but our duty to understand falls out of our duty to become more effective moral agents. This is, presumably, why this epistemological duty is not one that applies to all cognitive domains, but to the moral one in particular - only in the moral domain are we all subject to a duty of correct performance. Furthermore, the duty to understand is a positive, active duty; all agents have a duty to strive to increase their moral understanding. Presumably, this is because the duty to understand originates in a duty to act well, which is also a positive duty - our moral duties are not to pass some minimal moral prerequisite, but to actively strive to get it right very often.

If the end of being morally correct is capable of generating one active epistemic duty, it should be capable of generating others. There is, after all, more than one factor that contributes to getting it right. Understanding, certainly, contributes to getting it right, but another factor is having properly functioning cognitive faculties. Nickel is surely right that forsaking moral understanding will lead to getting it wrong often. But surely having an error in one's understanding process or a flaw in one's faculties of judgment would also lead to errors. And if the duty to get it right generates a duty to increase understanding, it should also generate a duty to discover and correct for malfunctioning cognitive abilities. And if Nickel is right that the goal of moral correctness generates positive duties, then, for the very same reason that we must actively seek to develop our moral understanding, we must also actively seek out other moral agents and use them to corroborate or dis corroborate our proper functioning. Both projects are substantial parts of the process of moral self-perfection.

My claims and Nickel's view fit so well because they are sensitive to the same fault-lines. When Nickel attends to Jones-style bias cases and allows for occasional deviations from moral understanding, he's taking into account the fact that we are fallible agents. Nickel and Jones are worried precisely about what I'm worried about: the need for a responsible moral agent to seek both understanding and engagement, and yet for a responsible moral agent to admit their own potential fallibilities, and move to make up for them with all tools available, including the moral judgments of others.

Understanding is a cognitive process. It is by its very nature aimed at getting things right. But the goal of getting it right also ought to inspire us to ferret out potential

problems and malfunctions in our own cognitive apparatus. The cognitive values of reliability and truth - of getting it right - call for many methods. Developing one's understanding is one such method, and cognitive self-assessment is another. If there is a positive duty for moral self-perfection, it ought to inspire us both to develop our understanding, and to use all means to check the reliability of our abilities and methods. Ignoring contrary testimony is surely a failure of my moral epistemic duties as failing to achieve moral understanding. In fact, if the duties of moral self-improvement are as pervasive and substantial as Nickel suggests, then it seems like we may have even have an active duty to regularly check our abilities, and so a duty to *seek out* contrary testimony and disagreement.

#### **Part IV. Alienation and objectivity**

If I'm right, then my conclusions might seem catastrophic for moral life. First, my claims seem to run contrary to the most laudable sort of moral commitment; it seems that I am calling for an end to moral heroism. Second, given the widespread occurrence of moral disagreement, it seems like my conclusions will force us to a sickeningly pervasive degree of self-doubt. Third, it might seem like I'm asking us to give up our moral commitments, and alienate ourselves from our deepest moral values.

First, it might seem like my conclusions might lead away from moral heroism and towards a tawdry sort of moral cowardice. After all, isn't the very *best* sort of person the one who stays true to their moral commitments in the face of opposition? Take, for



instance, an activist, dedicated to their cause, fighting for change against a stifling majority. The great moral heroes - those who opposed slavery, brought about women's suffrage, fought for the rights of immigrants - are those who fought with uncompromising hearts, who believed in their convictions to the fullest, who stood and fought with absolute commitment. It can seem that I am arguing against leading a morally wholehearted life.

I accept the charge that my arguments run against certain intuitions about what sort of person is the ideal moral hero. The epistemic analysis I've given shows that some of our intuitions about how moral life is supposed to go are simply unreasonable; intuitions, after all, are defeasible. The hyperconfident activist, who will admit no possibility of fallibility in her own thinking, is unreasonable. Any reasonable agent exposed to the social evidence of contemporary life must withdraw some degree of their confidence. But I do not think I am arguing against all forms activism, nor even moral commitment; what I am arguing against is an inflexible moral fanaticism, which regards itself as infallible. Remember, I do not argue for a suspension of our moral beliefs in the face of disagreement, only a withdrawal from unqualified self-confidence. It is still possible to act reasonably from one's qualified moral commitments. My argument will forbid only those actions which require complete or near-complete certainty.

It may help here for me to sketch a picture of where I hope this project will go. I have argued thus far that disagreement brings some form of moral self-doubt - a second-order doubt about the reliability of one's rational abilities and methodologies. I suspect that there are practical principles that will interact with this self-doubt, principles that

demand high degrees of confidence for action that intervenes in the lives of others against their will, especially when that intervention is radical or irrevocable. It might turn out, for example, that a committed vegetarian should work mightily to change the law, to speak and change the minds of others, to protest and march and the like, but that they should not, say, bomb meatpacking plants. In short, I hope that the principle of moral humility, combined with principles about restraint in action in the face of self-doubt, will produce something like an epistemically founded theory of tolerance, while leaving room for committed moral action.

Second, it seems like my conclusions will lead to pervasive moral doubt. For most of us, moral disagreements with peers are very common. Even if we attend only to the best cases of disagreement, the number of disagreements is still dizzying. Best-case disagreements occur over the permissibility of eating meat, about the permissibility of animal experimentation, about the permissibility of military intervention into non-democratic states, about the importance of ecological preservation compared to the importance of human jobs, about the right to suicide, about the degree of obligations one has to one's family, to one's state, to one's community.

This second charge I accept in its entirety. Lacking convincing, foundational moral theory, the only reasonable position given the presently available evidence is one with pervasive self-doubt and profound moral humility. In the face of disagreement, we cannot be entirely certain and entirely rational at the same. Our epistemic position is too poor at present, our rational resources too impoverished. This conclusion may be counter-intuitive, but, again, intuitions are defeasible.

Third, it may seem as if I'm asking agents to forsake their identity, to stand back from their core commitments, to alienate themselves from the values and moral beliefs that, in some sense, constitute their identity. But the use of moral testimony as I have described it is not an imposition from without. I've tried to show that the epistemic drive to self-doubt comes from within, not without. In each of the discussions of integrity, autonomy, and moral understanding, I showed that the reasons and rules an agent uses in coming to self-doubt are reasons and rules that any rational agent would be committed to - part of the commitment to accuracy that is partially constitutive of rationality. Thus, I do not accept the charge that my conclusions lead to alienation. They lead instead to a painful and complex process of self-analysis. The charge of alienation depends on the false presumption that our moral commitments are an essential part of our identity, but that our epistemic commitments are somehow alien. This cannot be the case for rational moral agents. Insofar as we take up rationality and its concomitant duties to accuracy, reliability, and truth, our epistemic commitments are also a crucial part of our identity. It is surely a difficult and painful process to step back from one's moral commitments to some degree, but this is a process fueled by other parts of one's rational commitments. Furthermore, insofar as our moral beliefs contain within themselves an aim of getting it right, they directly plug themselves in to all the norms of sound epistemic functioning.

Both the worries from autonomy and the worries from understanding are aimed at a particular sort of figure: an unthinking, obedient, passive agent who steps away from moral life, steps away from the responsibilities of being a rational, truth-seeking agent. But I am urging a use of testimony and a use of the disagreement which is the very

opposite of the unthinking, obedient, passive agent. In using moral disagreement as a method for self-assessment, we are acting from reasons as phenomenally personal, endorsed, comprehensible, and directly felt as any moral reasons: we are acting from epistemic reasons and standards that apply to any cognitive project. We are acting from our commitment to truth, accuracy, and objectivity. The fact that these reasons point us towards evidence in the external world does not make them alien; any process aimed at getting right some objective, mind-independent properties ought to tune us towards the external world. Reasoning in the space of objective cognition - of tracking truths independent of my phenomenal experience - must make use of such external evidence.

The motion of disagreement to self-doubt does lead to a painful, difficult friction. But this friction is not alienation - it is not a distancing of myself from my own reasons and commitments. It is a friction internal to my commitments, a friction that arises from the fact that I am committed to my moral beliefs as I feel them, but also committed to getting it right. I am committed both to the phenomena of moral beliefs and to their being aimed objective truth; these commitments can lead me in two different directions. This should be familiar, though painful territory. Our commitments often lead us into self-conflict, agony, and friction, especially when they run into the unforgiving world. Commitments to family, commitments to country, commitments to artistic and scientific projects - when these conflict, we don't call it alienation, we call it *tragedy*.

There is a difficulty to moral disagreement that is not there with empirical disagreement, I admit. This is because moral our beliefs are crucially *bivalent*. Our moral beliefs are both personal and cognitive. Our moral beliefs are important to us both as

expressions of ourselves, our identity, and important to us as trackers of truth. This is why moral beliefs are the flash-point for this conflict between different sets of reasons. Moral beliefs, unlike everyday empirical beliefs, are complexly entangled with different parts of reflective and practical life, and so make themselves subject to very distinct rule-sets. As parts of our personality and identity, they are subject to considerations of self-preservation and self-expression. It is important for self-preservation that I maintain and conserve key parts of my phenomenal existence, and my moral commitments are, indeed, key parts. It is important for my sincerity that I express, and act on my beliefs as I feel them. At the same time, our moral beliefs are important to us truth-trackers, and so they make themselves subject to the very impersonal rules and procedures of epistemic life. Any duties we have to be correct will call us not only to have an internal understanding of our moral beliefs, but to look outwards, to other people, to assess the reliability of our belief-formation process. But the impersonality of epistemic rules doesn't make them alien to the agent. They are a crucial part of the agent - the part of the agent that is rational, that seeks to attune their beliefs to the objective, external world. But though the evidence a rational agent uses may be impersonal, the essential motive is utterly personal - it is the motive of rationality.

The worry that we are alienating ourselves comes from an attention to the personal valence of moral beliefs. They are expressive of our personality, our identity. We, indeed, lose something of ourselves when we lose our moral beliefs. But the pressure against some of these beliefs, the pressure I'm describing, comes from a place that is also deeply embedded in any rational agent - the motive to get things right. In fact, this motive

is part of the moral life of a rational agent. I admit that the pervasive moral self-doubt that results attending to disagreement is difficult. It is painful. But it is not alienating.

## BIBLIOGRAPHY

- Ansombe, Elizabeth (1981). *Ethics, Religion and Politics* (Minneapolis: University of Minnesota Press).
- Baier, Annette (1986). "Trust and Antitrust," *Ethics* 96:2, 231-60.
- Baier, Annette (1992). "Trusting People," *Philosophical Perspectives* 6, 136-53.
- Burge, Tyler (1993). "Content Preservation," *Philosophical Review* 102:4, 457-88.
- Donagan, Alan (1977). *The Theory of Morality* (Chicago: University of Chicago Press).
- Dretske, Fred (2000). "Epistemic Rights without Epistemic Duties?" *Philosophy and Phenomenological Research* 60:3, 591-606.
- Enoch, David (2011). "Not Just a Truthometer: Taking Oneself Seriously (but not too seriously) in Cases of Peer Disagreement," *Mind* 119, 953-997.
- Foley, Richard (2001). *Intellectual Trust in Oneself and Others* (Cambridge: Cambridge University Press).
- Hardwig, John (1991). "The Role of Trust in Knowledge," *The Journal of Philosophy* 88:12, 693-708.
- Hopkins, Robert (2007). "What is Wrong With Moral Testimony?" *Philosophy and Phenomenological Research* 74:3, 611-34.
- Jones, Karen (1999). "Second-hand Moral Knowledge," *Journal of Philosophy* 96:2, 55-78.
- Kelly, Thomas (2006). "The Epistemic Significance of Disagreement," in Hawthorne and Szabo, eds., *Oxford Studies in Epistemology*, volume 1 (Oxford: Oxford University Press), 167- 196. Page references refer to online version from <http://www.princeton.edu/~tkelly/>.29
- Korsgaard, Christine (1996). *The Sources of Normativity* (Cambridge: Cambridge University Press).
- McGrath, Sarah (2011). "Skepticism about Moral Expertise as a Puzzle for Moral Realism," *Journal of Philosophy* 108:3, 111-37.
- Nickel, Philip (2001). "Moral Testimony and Its Authority," *Ethical Theory and Moral*

*Practice* 4:3, 253-266.

Plantinga, Alvin (1993). *Warrant: The Current Debate* (Oxford: Oxford University Press).

Pryor, James (2000). "The Skeptic and the Dogmatist," *Nous* 34, 517-49.

Quinn, Warren (1993). *Morality and Action*. (Cambridge: Cambridge University Press).

Sklar, Lawrence (1975). "Methodological Conservatism," *Philosophical Review* 84:3, 374-400.

Wallace, David Foster (2005). *Consider the Lobster* (New York: Little, Brown and Company).

Wolff, Robert Paul (1970). *In Defense of Anarchism*. (New York: Harper & Row).

Wright, Crispin (2004). "Warrant for Nothing (and Foundations for Free)?" *Aristotelian Society Supplementary Volume* 78:1, 167-212.