

UCLA

UCLA Electronic Theses and Dissertations

Title

Right Action and Integrity

Permalink

<https://escholarship.org/uc/item/80b2s4f4>

Author

Jennings, Justin

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Right Action and Integrity

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

by

Justin Jennings

2015

© Copyright by

Justin Jennings

2015

ABSTRACT OF THE DISSERTATION

Right Action and Integrity

by

Justin Jennings

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2015

Professor Barbara Herman, Chair

If there exists a right thing to do, why should the internal consistency of persons' attitudes necessarily matter, at all? Acting on one's best judgment or to taking the means one believes necessary to one's ends could lead one into incorrect action. Perhaps acting against one's best judgment or with inconsistent intentions could lead one into right behaviors or away from wrong ones. So why be consistent? Why not just be correct? Why should *akrasia* matter? Why not just right action? Why should rationality matter? Why not just reasons?

I try in my dissertation to answer these questions through an account of the structure of right action as such. I do not suggest one has reason in every case sufficient to make consistent action always right action. I argue instead acting on one's best judgment about what is right partially constitutes any instance of right action, at all. Right action, I argue, must be done because it is right, which suggests it must be done out of the attempt to do what is right. Yet one who tries to do what is right forms her best judgment about what is right and tries to act on it.

She may fail to do the right thing with consistency. Yet she cannot succeed in doing the right thing without it.

Chapter 1 formulates this account as a response to Niko Kolodny's critique of theories justifying internal consistency. Kolodny argues all the going theories either (1) entail one should act on one's beliefs merely because one has them, (2) give implausible reasons as to why one should be consistent, (3) require one to act for the sake of consistency rather than correctness, or else (4) do not explain why one's beliefs should guide one's actions, at all. An accurate theory must avoid all these unacceptable results. This account of right action avoids them by suggesting not the sufficiency but the necessity of consistent action to right action. One must form and act on one's best judgment as a constitutive part of carrying out any particular right action as such.

Chapter 2 argues for the sufficiency of the above account of right action to ordinary uses of the concept by responding to an objection Nomy Arpaly raises against accounts of right action as entailing action on best judgment. She argues no such view can account for the actions of the titular character from Mark Twain's *The Adventures of Huckleberry Finn*. Huck, being from the antebellum American South, believes he should report his friend Jim to the authorities because Jim recently escaped enslavement. But after their journey down the Mississippi River together, Huck finds he cannot go through with it and instead helps Jim escape. Arpaly claims Huck acts against his best judgment and yet still acts rightly. I argue Huck does act rightly but acts on a non-deliberative form of judgment on which persons appear to act over the vast majority of their lives. Huck sees Jim's humanity and inchoately knows he cannot turn him in. Persons who "act against their best judgment" as Huck does seem to act rightly where those merely causally overpowered by emotion or desire do not.

Chapter 3 presents two accounts of the Instrumental Principle. Wrong ends do not make it right to take the means one believes necessary to them. So what does one mean in saying, “I don’t think you should be a musician, but if you’re intending to be one, shouldn’t you be practicing?” What is this “should”? I give two accounts. First, persons cannot achieve ends if they do not intend means they believe necessary to achieving them. One must achieve ends intentionally to count as achieving them. But intentionally achieving ends entails intending the means one views as necessary to achieving them. Instances of instrumental inconsistency thus give evidence of one’s general unreliability for right action. Second, actually correct ends seem to count in favor of the means necessary to achieve them. One who believes an end to be correct but not the means she views as necessary to it thus has inconsistent beliefs. Someone trying to do what is right necessarily pursues an accurate view of the facts and act on it. But an inconsistent view cannot be accurate. Instances of instrumental inconsistency thus give evidence a person lacks the intention to do what is right, which right action as such requires.

Chapter 4 attempts to explain the relationship Bernard Williams identifies between right action and integrity. Pure consequentialism views the right action as in every case the action producing the maximally good state of affairs. Williams argues such views of right action take inadequate account of the cost such actions impose on persons’ commitments. For instance, one usually takes it one should look after one’s family even though seeing to the welfare of strangers might produce greater aggregate happiness in the world. Intuitively, Williams argues, one’s integrity, one’s fidelity to one’s commitments, partially determines when one acts rightly and what it is right for one to do. I suggest that under a conception of integrity closer to ordinary language this argument holds not merely intuitively but deductively. Under the above account, forming and trying to act on her best judgment about what is right partially constitutes right

action as such. But integrity ordinarily means forming and trying to act on one's best judgment about what is right. So, on the one hand, acting against one's best judgment constitutes failed right action in every particular case. On the other hand, action constitutively requiring one not to form or act on one's best judgment cannot be right action, at all. Right action's structure rules out pure consequentialism *a priori* as one of its possible contents. Pure consequentialism cannot give the content of right action inasmuch as right action has the structure it ordinarily seems to have, that is, if one must do the right thing because it is right.

The dissertation of Justin Jennings is approved

Joshua F. Dienstag

Gavin Lawrence

Barbara Herman, Committee Chair

University of California, Los Angeles

2015

for Gary and Debra Jennings, my parents

TABLE OF CONTENTS

Abstract of the Dissertation	ii
Committee	vi
Dedication	vii
Acknowledgements	ix
Vita	xi
INTRODUCTION	1
CHAPTER 1: Right Action, Second-Order Intention, and Internal Consistency	12
CHAPTER 2: Right Action and the Scope of Judgment	41
CHAPTER 3: Two Accounts of the Instrumental Principle	66
CHAPTER 4: Right Action and Integrity	105
Bibliography	137

ACKNOWLEDGEMENTS

Thanks are due first to Barbara Herman. This dissertation would not have attained its present form without her guidance, the precision of her critique, the influence of her thought, the subtlety of her readings of philosophical history, or her ability to reframe arguments so that one sees even one's own thought anew. She has been the principal guide of my doctoral work from the very beginning. I am thoroughly indebted to her.

Thanks are likewise due to Seana Shiffrin. Her supervision during my Proposition year marked a crucial stage in the development of my scholarly writing and her influence on every aspect of my work has carried throughout this dissertation's composition. I owe my knowledge of the craft and practice of the academic profession substantially to her advice and her help.

A.J. Julius deserves especial thanks for his role in this dissertation's production. Not every theorist would give his time and his thought to another laboring on questions similar to his own. Fewer would actively, generously, and continuously aid the other in that labor. A.J. did. I cannot thank him enough.

Gavin Lawrence's constant readiness to rethink things from their foundations and to explore radically new approaches has made him one of my most valued interlocutors. His lectures and writings on Aristotle altered utterly my understanding of that philosopher and helped spark some thoughts that made their way to this dissertation. I am very grateful to him.

I owe great thanks as well to Pamela Hieronymi. Her utterly original work, her insistence on plain speaking, and the rigor of her thought all helped shape this dissertation and served as some of its models. Her aid and counsel has been crucial to my development as a thinker and to the progress of my doctoral study.

Joshua Foa Dienstag of the Department of Political Science at UCLA gave excellent advice in determining the direction this work should take and in keeping it true to insights from some of its historical antecedents. I am very grateful for his help in its production.

For their help prior to my entrance into doctoral work I would also like to thank Warren Goldfarb, Christine M. Korsgaard, and T.M. Scanlon of the Department of Philosophy at Harvard, as well as Richard Heck, now of the Department of Philosophy at Brown. I would have been unable to enter into doctoral study in philosophy without their instruction and aid.

I also would like to thank Charles M. Lewis of the Department of Philosophy at Wake Forest, the foremost model I have known of philosopher as Educator. I am one of a long line of students inspired to a life of concerned philosophical reflection by his teaching and his example.

Too many people have helped my thinking over the years to enumerate them all. An inadequate alphabetical list would be this: Ingrid Albrecht, Gregory Antill, David Bordeaux, Denis Buehler, Tyler Burge, John Carriero, Lee Ann Chae, Lindsey Chambers, Ryan Doerfler, Molly Faulkner-Bond, Derek Elkins, Ashley Feinsinger, Darren Frey, David Friedell, Laura Gillespie, Jonathan Gingerich, Peter E. Gordon, Mark Greenberg, Louis-Philippe Hodgson, Andrew Hsu, Brian Hutler, Gabrielle Jackson, Andrew Jewell, Rachel Johnson, Rachel Kaminer, Ben Kesling, Michael Kessler, Sari Kisilevsky, Scott Klein, Philip Kuberski, Kevin Lande, Michael Lopez, Dorothea Lotter, Allen Mandelbaum, Harvey Mansfield, Adam Masters, Aaron Mead, Eliot Michaelson, David Moore, Richard Moran, Calvin Normore, Kimberley Patton, Alexander Patsaouras, David Plunkett, Joel Rasmussen, Pollyanna Rhee, Lauren Schaeffer, Tripp Smith, Gavin Stephens, my students, Somya Tariyal, Sabine Tsuruda, Robert Utley, Drew Walker, Tamar Weber, Daniel H. Weiss, Stephen White, and Aaron Winter.

Thank you above all to my family. Everything I am and do I owe to you.

VITA

2002 B.A. Philosophy; B.A. English
Wake Forest University
Winston-Salem, North Carolina

2005 Master of Divinity
Harvard Divinity School
Cambridge, Massachusetts

2005-2007 Teaching Assistant
Harvard University
Cambridge, Massachusetts

2008-2015 Teaching Assistant
Department of Philosophy
University of California, Los Angeles

2010 M.A. Philosophy
Department of Philosophy
University of California, Los Angeles

2013- Instructor
Department of Philosophy
University of California, Los Angeles

INTRODUCTION

Why make your best judgment about what is right and try to act on it if it sometimes might lead you wrong? Why be careful if laxity might sometimes produce a better result? Why should it matter whether persons pursue their own values, so long as they behave correctly? Why have integrity if you could do the right thing without it? Maybe one usually goes the right way in thinking and acting with probity, rigor, courage, or fidelity. But must one *always* do so? Maybe sometimes their opposites would lead one the right way. Maybe sometimes they could even constitute right action in themselves. One's best efforts to see clearly and do the right thing may go awry. So why think one must always make them?

I do not argue in this dissertation that action with integrity is always right action, always something one has sufficient reason to do. I argue forming and trying to act on one's best judgment about what is right partially constitutes all right action, as such. Action consistent with one's best judgment does not suffice for right action in every case. Yet right action in every case necessitates that persons act on their best judgment in doing it. One must make and try to live up to one's best judgment about what is right because that is what one does who *tries* to do what is right. But one cannot do the right thing at all unless one tries. So one must always try.

Ordinary language seems to suggest one does not do the right thing unless one does it *because* it is right. Good events may occur through one's behaviors, yet one cannot be credited with acting rightly in bringing those events about unless one brought them about because it was right to do so. One cannot do the right thing by accident.

Yet the nature of this "because" is not immediately obvious. It seems clearly not to be the "because" one uses in the sciences to refer to the outcomes of merely causal exchanges. One

would not seem to act rightly if one performed the appropriate behaviors as purely causal reactions to those actions' correctness. Though one performed those behaviors "because" they were right, it would not seem to credit to one, oneself, that one did the right thing. One seems instead to do something "because" it is right in the way one goes to the hospital "because" one is sick or goes to work "because" one needs to feed one's family. One acts in respect of some state of affairs. One actively undertakes an action on the basis of something's being the case. In these two cases, one acts in respect of one's illness and the illness-alleviating aspects of going to the hospital or in respect of the need of money to acquire food and the money-generating aspects of going to work. One acts on the basis of the fact of an action's correctness when one carries out that action because it is right.

In performing right actions, then, one must actively undertake those actions on the basis of their correctness. One must, that is, act out of a second-order intention to do what is right, carrying out actions in light of whatever reasons suffice to make them right. One must try to do what is right not because trying suffices for right action. One must try to do what is right because trying to do the right thing partially constitutes right action. Inasmuch as one always must do what is right, one always must act out of the intention to do what is right.

The argument that right action constitutively entails action on best judgment proceeds by analyzing this intention. In this second-order intention, one sets out to do the right thing. That means one sets out to act on the basis of actions' correctness. But what does one set out to do when one sets out to do that?

Acting on the basis of a fact seems constitutively to require one to recognize that fact. One cannot obey a command if one does not hear the command or does not hear it *as* a command even if one happens to perform the action commanded on some other basis. One cannot follow a

set of instructions if one does not know of their existence or does not understand them. One cannot act on the basis of a fact one does not recognize any more than one can believe something on the basis of evidence contained in books one has not read or experiences one has not had. The intention to do the right thing thus seems to have as one of its constituent parts the intention to come to a correct view about what is right.

Yet intending to come to a correct view about what is right is not a simple intention, either. One does not seem to have that intention if one submits to having whatever views one merely happens to have or is causally impelled to have by one's psychology, society, or position in history. The sequence of events and causation do not necessarily track the truth. Nor does one seem to intend to come to a correct view if one intentionally thinks less carefully about what is right than one could. One who intentionally reads a veridical science journal uncarefully intentionally submits to having less accurate beliefs than she could have with more or better effort. A person with the intention to have an accurate view, inasmuch as she retains that intention, cannot intentionally submit to that. One who commits to recognize the facts seems thereby to commit to maximally overcoming her accidental circumstances and limitations and following the evidence with maximal accuracy.

Yet persons differ in their ability to overcome their circumstances and in the character of their most careful reasoning. Though one may intend to follow the evidence, one cannot follow it perfectly by merely intending to do so. Causal interferences, especially one's own nature as an agent reasoning in space and time, make up fixed parts of the context within which one has to work out the facts. One is not immediately delivered from that context just by intending to see things objectively. It seems one who tries to come to an accurate view of the facts thereby works to the best of her ability to overcome her limitations and follow the evidence to the facts. She

forms her best judgment about what is right. “Best judgment” here does not mean the way one judges when not much hangs on the issue or when the evidence is inconclusive. Rather, it means the judgment that is literally the best one is capable of producing. A person forming her best judgment in this sense does not only try to do her best. She tries her best to get it right.

To an outside observer, a person forming her best judgment generates a mental state that may be true or false. Yet from the perspective of the person forming her best judgment, she comes to view of the facts themselves that is accurate, as far as she is able to tell. Having done everything she could to get it right, she cannot see around it. It is the way things are, as far as she can make them out. The specific content of this view will differ between persons based on their abilities. For instance, an expert scientist and an novice investigator may read the same science text to determine the facts about some subject matter and arrive at very different views as of those facts through that reading. Yet if both work to the extent of their abilities to understand and follow the evidence, the relation each person’s judgment would bear to her intentions, her abilities, and the evidence she recognizes would remain the same. No less with the expert than with the novice, it would be her best judgment, would follow the evidence she recognizes as closely as her abilities allow. Neither person will think of it in the first instance as a mental state. It will be, as far as she can tell, an accurate view of reality itself.

So when a person sets out to do the right thing, she sets out to arrive at an accurate view of the facts about what is right and to act on those facts. But in pursuit of that project, she forms her best judgment about what is right and tries to act on the content of that judgment. She does not set out to form or to act on her best judgment out of concern about the quality or consistency of her mental states as such. She will do so because she intends to act in consistency with what is right, what is *actually* right. Forming and trying to act on one’s best judgment make up parts

of that intention. They are not what she has in mind but rather the background mental processes through which she pursues what she does have in mind, which is doing the right thing.

This structure explains the universal “should” of internal consistency as follows. There may or may not be particular reason to work to make one’s mental states and actions consistent. Nonetheless every person who acts rightly in any given case necessarily works toward a state in which her mental states and actions are consistent. To intend to do the right thing is to intend to make one’s actions and intentions consistent with the facts about what is right. But making one’s actions and intentions consistent with the facts constitutively necessitates bringing one’s actions and intentions into consistency with one’s best judgment as to the facts. One works toward right action rather than consistency. But right action itself necessitates consistency. One may still fail to do what is right when one pursues right action for its own sake. But one must pursue right action to *succeed* in right action, at all.

So the universality of the “should” of internal consistency does not express something it is always right for persons to do. One should try in every case act or intend consistent with one’s best judgment about what is right because that is what a person with the second-order intention constitutively *would* try to do. Yet one must have the intention to right action intention to carry out any right action for its own sake, at all. A case of internal inconsistency may arise because of several different factors. One may not fully understand the content of one’s beliefs, may be unaware of the inconsistency between one’s actions or intentions and one’s views, may lack the time to think through all of one’s convictions, may have attitudes one genuinely cannot change. Yet one explanation always available is that one is not trying to do the right thing. But because one must always do the right thing, one must always try to do the right thing, as well.

Chapter 1 formulates this account as a way of responding to Niko Kolodny's critique of the justification of internal consistency forwarded in his "Why Be Rational?" Kolodny showed theories attempting to ground internal consistency in some particular reason or value succumb to one of four pitfalls. (1) They entail "bootstrapping," or that one should act on one's beliefs about what is right just because one happens to have them. (2) The reasons they give are implausible. (3) They require persons to act or intend in ways that are clearly irrational in, for instance, guiding their actions by consistency as such rather than by what they have reason to do. Or else, (4) they fail to explain why one's beliefs should guide one's actions, at all.

The above account avoids all these pitfalls because it does not suggest consistent action is right or reason-favored action. (1) One is in no way justified in acting on one's beliefs, merely as such. (2) No reasons at all are posited in consistency's favor. (3) Trying to do the right thing in fact constitutively requires one *not* to act for consistency's sake. One does not act out of the second-order intention, at all, if one undertakes one's actions on the basis of something other than their correctness. But one must act out of the second-order intention to do what is right in any case, at all. Finally, (4) one's beliefs must guide one's actions not because it is right to act on them but because acting on one's best judgment is a necessary condition of right action as such. One must form and act on one's best judgment about what is right not as right action but as part of the process of right action. Thus the very content of the ordinary concept of right action seems to provide an account of why one "should" act with internal consistency.

Yet where the above account of right action seems to follow from right action's ordinary concept, a question still remains as to the account's sufficiency, whether it tracks ordinary usage in particular cases. Chapter 2 suggests its sufficiency by responding to an argument forwarded by Nomy Arpaly whereby right action cannot possibly entail action on best judgment. Arpaly

brings forth the titular character from Mark Twain's *The Adventures of Huckleberry Finn* as evidence. Huck, being from the antebellum American South, believes he should turn his friend over Jim to the authorities because Jim has recently escaped enslavement. Yet having become close friends with Jim on their journey down the Mississippi River, Huck finds he cannot do it. Instead Huck does the actual right thing and helps Jim escape. Arpaly claims Huck acts against his best judgment yet still acts rightly, thus right action cannot entail action on best judgment.

I argue this interpretation of Huck's case lives off an unfounded assumption about the scope of judgment. Persons spend the vast majority of their lives acting without deliberation. Yet their non-deliberative action exhibits a control, complexity, and articulation inadequately accounted for in terms of merely causal mechanisms. It seems they act instead on a non-deliberative form of judgment about what is right. Huck's struggle appears not to be between the causal force of his emotions and reason, at all. Rather, he struggles between his deliberative convictions about slavery and the judgment of his own non-deliberative experience that Jim is a human being and that it is right to help him escape. In fact, on a survey of cases in which persons "act against their best judgment" as Huck does, it seems one would view them as acting rightly, yet where emotions or desires overcome a person *merely* causally, one does not. With the scope of judgment widened to capture actual human experience, the above account of right action thus seems wholly sufficient to one's ordinary use of that concept and to give an interpretation of Huck's situation truer to Twain's text, as well.

Chapter 3 employs the structure of right action to suggest two accounts of the Instrumental Principle. The Principle states that one in some sense "should" intend the means one views as necessary to one's ends in every case. For instance, one can say that a person "should" practice an instrument if she intends to be a musician even if one does not believe being

a musician is the right thing for her to do. Yet it seems wrong to think having incorrect ends somehow makes it right to take the means one believes necessary to those ends. It cannot be right to do what is necessary to achieve one's ends just because one has them. So what is the sense of this "should"?

First, I argue, right action itself constitutively requires persons to act and intend with instrumental consistency. One cannot achieve an end if one does not carry out means necessary and sufficient to achieving that end. One thus cannot achieve an end if one declines to take means one views as necessary to achieving it. Yet if one cannot intentionally produce an end, one cannot intentionally produce that end because it is right, either. Particular instances of instrumental inconsistency, regardless of the right or wrong of one's end, thus give evidence of one's general unreliability for right action. When the Instrumental Principle is uttered in this sense, it expresses the necessity of being willing and able to do the right thing, not merely in intention but in execution, as well. Second, I argue that if actually correct ends count in favor of the actual means necessary to bring them about, a person believes inconsistently who believes an end is right without believing it to count in favor of the means one believes necessary to it. Someone trying to do what is right would try to resolve this inconsistency in her beliefs, to pursue the end, and to take the end's necessary means. One seems to presume in everyday discourse that one may appropriately question persons as to their actions' justification. This seems founded on a further presumption that persons generally pursue ends they view as right. Within that context, instances of instrumental inconsistency thus provide evidence a person lacks the intention to do what is right, which all right action constitutively requires. Used in this sense, the Instrumental Principle expresses that requirement.

Finally, Chapter 4 tries to explain the relationship identified by Bernard Williams between right action and a person's integrity. Williams argues no pure consequentialism can give the content of right action. Consequentialisms require persons in every case to produce the maximally good state of affairs. The qualitative differences between courses of action can make no difference to whether it is right to perform them. So though the cost to one's personal commitments may count *somewhat* against a certain course of action, if the calculus of values still weighs in favor of one's carrying out that course of action, one must carry it out regardless of that cost. One's integrity, one's fidelity to one's own commitments, can play no conditioning role in determining when one acts rightly or what it is right for one to do. Intuitively, Williams argues, a correct ethical theory allows for integrity to play that role. Consequentialisms' wholly quantitative comparison of actions' disallow them from making sense of that role. Thus, Williams argues, pure consequentialisms cannot give the content of right action.

I suggest that under an adequate conception of integrity, this argument is not just intuitive but deductive. Under the account above, right action must be performed as the outcome of action on best judgment. But trying one's best to determine and do the right thing, that is, forming and trying to act on one's best judgment about what is right, seems to be what is meant in ordinary language by "integrity." This conception of integrity seems also to draw evaluatively relevant distinctions between different ways persons may act for or against their personal commitments. Integrity thus seems partially to constitute right action as such.

This structure appears to provide systematic accounts of both aspects of the role Williams saw integrity to have in conditioning the content of right action. On the one hand, a person cannot count as acting rightly in any particular case in which, as far as she can tell, the action she performs is incorrect. One might perform the right behaviors or might bring about good states of

affairs in so doing. Yet in acting against one's best judgment one could not do the right thing because one could not do it in light of its correctness. On the other, action constitutively requiring one *not* to do one's best to determine and do what is right cannot be right action, at all. Pursuing the second-order intention to right action constitutively requires forming and trying to act on one's best judgment about what is right. Giving up either of the latter projects thus entails giving up the former. Yet one cannot perform right action by giving up the second-order intention to right action. One cannot succeed in doing the right thing by forfeiting the attempt. But declining either to form or to act on one's best judgment entails that forfeit. Pure consequentialism thus cannot give the content of right action if right action has the structure ordinary language seems to suggest it has.

One does not do what is right unless one does it for its own sake. One does not do what is right for its own sake unless one does it out of the attempt to do what is right. So it may not be possible to say *a priori* what is right for every case. Yet one can say two things *a priori*, if nothing else. First, giving up trying to do what is right, what entails giving it up, and what giving it up entails cannot be right. It cannot be right because, second, whatever may be right, one must do it out of one's intention to do what is right. Whatever is right, one must try to do the right thing. One must try as best one can.

Before beginning in earnest, I should make the following note. For methodological reasons connected to the explanation of consistency principles, I must often speak about the evaluation of the actions of others as though it were my primary concern. Though that issue is of vital importance, my foremost concern lies in the question, "How should I live?", in the necessity of asking it, and in what pursuing and living out its answer demands of one. These concerns produced most of the thoughts expressed in this dissertation, especially those at its foundation.

Yet I would like to point out that the principles explained through the examination of others' actions seem to have an application to one's own attempt to live as one should, as well. One must often examine oneself in the course of that project, must try to determine one's own character, one's own capacities, what one's motives really are. It seems one may apply these principles to oneself *as to another* without loss of sense in that process of self-examination.

CHAPTER 1

Right Action, Second-Order Intention, and Internal Consistency

1. Introduction

Say there is a right thing to do. The existence of correct answers about what is right would render it mysterious why one should care whether persons live up to their own beliefs about what is right if one does not share those beliefs. Take Niko Kolodny's example of someone who believes in God and believes that God made everyone equal but still does or intends racist things.¹ It seems one could and would want to say to such a person, "If you believe God made everyone equal, shouldn't you treat everyone with equal respect?" One would want to say this even if one believes in no deities. But what is this "should?" Or say one's friend belongs to a religious sect that requires her to give the sect half her income every month, yet she only gives a quarter. One could say to her, "Look, I do not agree with this sect's views, at all. But if you really believe in it, oughtn't you to be giving half?" and do so with sense. But what is this sense? How could it make sense in every case to say that a person "should" or "ought" to do what she thinks is right, even if one thinks she is wrong?

One could understand this as a question about what reason or value makes it right to take on the second-order intention to make one's beliefs and intentions consistent in every case. Niko Kolodny has shown this question to have little hope of a positive answer. In his "Why Be Rational?" Kolodny shows all the going theories of this kind fail one of three tests: either (1) they entail "bootstrapping," that attitudes are self-justifying, (2) they suggest implausible reasons, (3) they entail that one must reason on the basis of consistency requirements, which one neither does nor in some cases *could* do, or (4) they fail to account for the genuinely intention-guiding role a person's own beliefs would seem to have under any adequate theory.

¹ Niko Kolodny. "Why Be Rational?" *Mind* (July 2005) 144: 509-563, p. 554.

I dispute none of these criticisms. Rather, it seems the question as to why one must take on this second-order project has been wrongly interpreted. I argue that persons must try to act on their own best judgment about what is right not because consistent action is right action. Rather, they must try to act on their best judgment about what is right because internal consistency is necessary to all right action, as such.

Ordinary language seems to suggest one does not act rightly unless one does what is right *because* it is right and that one does what is right because it is right if one is *trying* to do what is right for its own sake. To do the right thing, one must act out of the second-order intention to do what is right. If this is correct, I argue, then trying to do the right thing is to form one's best judgment about what is right and try to act on it. Where one succeeds even by one's own lights, one's beliefs and intentions necessarily will be consistent. But they will be consistent because one sought not consistency but right action. Within this structure, instances of first-order inconsistency thus give defeasible evidence one is not trying to do the right thing. Yet because all right action must be done out of that attempt, it is problematic in every case for a person not to make it. So because inconsistency is defeasible evidence of that intention's absence, inconsistency is defeasibly problematic in every case, as well.

First, to fix ideas, I set forth the sorts of cases in question. Second, I lay out more fully the account of right action and the structure of the second-order intention. Third, I show how this account makes all instances of inconsistency defeasibly problematic. Fourth, I set out Kolodny's tests for a successful theory. Fifth, I show how this account passes those tests.

2. Cases

Kolodny observes there are in fact *two* ways a person's attitudes may be internally inconsistent. The first kind most concerns him. As in the above case, a religious racist believes in God and that God made everyone equal and yet still intends racist things. Her beliefs about what is right recommend or proscribe some action but she does not intend in accordance with those beliefs. It is a simple relationship of incongruity between her mental states.

Yet Kolodny points out that one can also be inconsistent between one's beliefs and intentions because one has arrived at one's intentions in the wrong way. Consistency requires one to intend what one intends *because* it is, one believes, right. One is not consistent if one arrives at even a congruent intention arbitrarily or on some basis other than the intention's correctness. An instance of this form of inconsistency might be a certain kind of good-natured egalitarian, who believes everyone is equal, that human equality entails that everyone deserves to be treated with equal dignity, and whose intentions are in fact congruous with those beliefs. Yet she does not come to these intentions *because*, so far as she can tell, it is right to act in such a way. Deep down she comes to them out of her drive to be well-adjusted socially, and that is just what is believed and done or intended in her community. Her attitudes are congruous but not related to one another in the right way to constitute genuine internal consistency.

Both forms of inconsistency would seem problematic in some way. Yet this appearance seems to change on learning further facts. The person never had the time or the opportunity to sit and think through all the implications of her beliefs. Or she does not fully understand the contents of those beliefs. Or she has urgent responsibilities precluding for the moment her engaging on a process of self-examination. In light of such facts, persons' inconsistency with their beliefs about what is right appears less problematic. The inconsistency seems to signal

something amiss and yet certain further facts mitigate that signal. The person's actions may still appear as incorrect. But one does not see *her* the same way.

Now suppose the following. A person intends or acts inconsistently with what she believes is right, fully aware of her inconsistency, fully understanding it, with a corrigible psychology, and with time available to consider and change her attitudes to fit what she sees as right. Yet despite these favorable circumstances she makes no effort to live up to her own views or to change her inner relation to her actions. She persists knowingly in her inconsistency. Cases such as these would seem to appear problematic regardless of the circumstances and seem to alert the theorist to a third kind of inconsistency.

It seems persons may not concern themselves in every case with doing what the reasons favor, with doing what is right. They do or intend what is useful to their career goals, gives them a feeling of moral superiority over others, appeals to them in the moment, or what satisfies some other criterion other than correctness, or else may even intend arbitrarily. To intend and act in this way would not be mere immoral or irrational action, but *amoral*, *arational* action. Such a person would intend inconsistently not merely in having incongruous attitudes or arriving at some attitudes in the wrong way but in declining to guide her actions on the basis of what is right, at all. Theorists often imagine such characters as wantons and libertines. Yet it is not at all necessary for the person's actions, intentions, or beliefs to diverge from what would be, in a certain sense, the correct ones to perform or to have. In all her outward and even inward behaviors she may appear the very image of rectitude, doing, intending and thinking what one is *supposed* to do, intend, and think. Yet those behaviors appear as hollow, as wrong in a way deeper than simple incorrectness, when one discovers the person's true motivation to merely further her career, feel morally superior, or gratify herself, or even finds out nothing matters to

the person in a determining way at all. These correct behaviors would not seem to count as a person's doing the right thing. The *person* would not do it. She was not *trying* to do it, at all.

It seems, however, that the sense of something problematic in cases where persons do not try to do what is right would not dissipate on learning, for instance, of the person's incomplete understanding or lack of time for reflection. Say the religious racist or good-natured egalitarian only cares about maintaining her comfort or social belonging rather than doing what is right. If these things were so, her acquisition of a complete understanding of her beliefs or the time to reflect on them would not necessarily change her response to circumstances as it would for someone trying to do what is right. Her inconsistency still would not appear to her as relevant, as a basis on which to change her actions or her inner orientation to those actions, except inasmuch as it would help her maintain her comfort or fit better with her community than to persist in inconsistency. Though she may she may give outward signs of trying to change her racist actions or of undertaking a process of self-examination, it would appear she does these things, say, only to avoid social opprobrium and so ensure her maximum comfort or to assure her community of her genuine commitment. Two things, however, would seem to change this appearance. On the one hand, the observer realizes the person really has begun to try to do what is right. Or, far more commonly, the observer realizes she was wrong about the person in question all along. The seemingly amoral or arational person *was* trying to do the right thing, just, say, with false or incompletely understood beliefs, a psychological inability to change, or within a narrow or distorted value system she was unable to see around.

So it seems three aspects of internal consistency are already familiar: (1) first-order inconsistency is universally but defeasibly problematic, yet (2) the absence of the second-order intention to do what is right is universally *indefeasibly* problematic. Lastly, (3) one explanation

of a person's first-order inconsistency in either first-order way is that she lacks the second-order intention. So the position argued in this chapter seems familiar. I will now try and suggest that at least in this case what is familiar is also correct.

3. Right Action and the Second-Order Intention

Ordinary language and practice seem to suggest right action must be done *because* it is right, that to count as doing what one's reasons require, one must act *for* those reasons. If a good event is caused by a person's actions, and yet it happened as it were *by accident*, that event does not count as right action, even if it may also have been the correct behavior to perform. Putting money in the hands of a needy person in the thought she is selling something one wants to buy does not count as performing the right action of giving to the needy. So neither does happening to drop it in her proximity. But what is the sense of this "because?" In what sense must an action's correctness be non-accidental to one's carrying it out for one to act rightly in doing so?

It seems clearly not to be the sense employed in the sciences whereby a purely causal force or mechanism brings a person passively to do the correct action. Such causes may produce unbroken series of good behaviors or good events in persons. Yet they would not seem to be instances of the person herself doing the right thing. It would be as though correct behaviors occurred *through* her, but she herself would not be the one doing them, not the one acting rightly.² But what then must be in place for a person to do the right thing because it is right?

It seems the sense of "because" of doing the right thing because it is right is the same one uses in any other case in which one looks for people's motivations. One looks for the facts on the

² I certainly do not mean to suggest that persons only act rightly, having *deliberated* over and *decided* what is right. That entailment would require the assumption that all non-deliberative action must be the outcome of merely causal forces or mechanisms. This assumption seems to have very little support in the data of non-deliberative action it is meant to explain. In short, one should never confuse *reaction* with *response*. For more on this, see Chapter 2, "Right Action and the Scope of Judgment."

basis of which a person determines herself to act the way she does. When one jumps in the lake “because” one wants to go for a swim, one undertakes to do so in light of the facts of the situation: one is on a bridge over some water, say, and one wants to. Whereas when one falls into the lake “because” there was a hole there one did not act in light of the fact of the hole’s existence. Rather, a state of affairs forces one passively into in the lake. One did not do anything on the basis of the facts of the case. Rather, the facts of the case did something *to* one that happened to have the same or a similar outcome as if one jumped in. Unless some deep naturalist materialism is true, the lake and the hole in the bridge exert no causal force on one when one jumps in the lake. One acts, oneself, *with respect to* the facts of the lake, the hole, and one’s desire to swim. In light of the circumstances, one determined to jump in and one did.

So if the “because” of right action is this active sense, right action requires one to intentionally *respond* to an action’s correctness with the intention to do it rather than to causally *react* to it. One must actively undertake actions on the basis of their correctness. But actively undertaking actions on the basis of their correctness seems, in turn, to entail that one must form the intention to do them out of an intention to do what is right. To do the right thing, one must be *trying* to do the right thing. One talks this way about right action all the time.

If correct, this account shows why a person’s declining to try to do what is right would always be problematic. It would not be because having that intention is necessarily correct in itself. Rather, it would be because that intention constitutes a necessary condition of all right action as such. As in the cases set out before, this is already how persons appear who are, for instance, fundamentally manipulative, self-righteous, or hedonistic. In doing the right actions but doing them because of the manipulative, self-righteousness-supporting, or pleasure-producing qualities of those actions, their actions appear as a kind of pantomime or twisted version of

correct behaviors, not as genuine right action. I now want to suggest this structure explains why the internal inconsistency of beliefs and actions is always defeasibly problematic.

4. The Second-Order Intention and First-Order Consistency

A person who intends does not merely wish or suppose something to happen. She has already actively committed herself to carrying out an action. The intention is the totality of her contribution to doing so. If things do not turn out the way she intends, it can only be because the world does not cooperate with her intention, not because of anything she herself fails to do. So the structure of the project, what she has it in mind to do, makes up the intention's content.

In trying to explain the "should" of internal consistency, I will analyze the intention to do the right thing for its own sake. In this way I argue that a person genuinely trying to do what is right has necessarily *already* committed to bringing a state of consistency to her beliefs and intentions. Yet she has committed to bringing about this consistency not because she has it in mind to be consistent for its own sake. A person with this second-order intention concerns herself only with doing what actually *is* right, not what she merely believes to be right. However, doing what is actually right entails acting consistently with one's best judgment about what is right. Internal consistency is a necessary third-personal byproduct of right action and of certain of its component parts. One who tries to do the right thing thereby tries to bring a state of consistency to her attitudes. But one who does not act consistently does not act rightly, either.

First I explain what it is to try to act "because" of some fact. Then I explain what it is to try to recognize facts and respond to them in action. Finally, I show how a person's efforts in this project result when successful in the consistency of her attitudes. This then will explain why a

person's inconsistency is defeasible evidence in every case of her lacking the second-order intention. In brief, it is because the intention's absence is *indefeasibly* problematic in every case.

4.1 "Because" and Recognition

In the second-order intention one sets out to do the right thing in light of its being right, to do what one has reason to do on the basis of having that reason. But what is necessary to do this? What conditions must be in place to guide one's actions by some fact?

It would be hard to understand how one could stop on the basis of a stop sign if one were unaware of the stop sign's presence. One may stop at the place where it stands by accident, say because one thought one saw a deer jump across the road or by causal force, say by running into the sign. Yet in such cases it appears one stops on a basis other than the stop sign or on no basis at all. In these cases, the person does not respond to the stop sign *as* a stop sign, but either responds to something else or reacts to a causal factor within the environment. So it seems that to act or intend in respect of a fact one must recognize that fact. One cannot do or intend the right thing because it is right unless one recognizes the fact that the right thing is right.³ Trying to do the right thing is to try to act in *recognition* of the correctness of right actions.

At first, this may seem to make no progress toward accounting for internal consistency requirements. When one tries to do the right thing, one does not think about one's mental states at all. When trying to determine the right color to paint a house, one looks to the house, to the landscape, to the colors available. When trying to determine how to respond to a car accident outside, one looks to see if anyone is hurt, if someone needs to be helped out of the car, if there are any oncoming cars that need to be stopped, and so on. One does not – and indeed *should* not

³ Nomy Arpaly offers significant challenges to this view in her *Unprincipled Virtue*, which I try to address in Chapter 2, "Right Action and the Scope of Judgment."

– consult one’s mental states. It would be a symptom of extreme narcissism or sociopathy to think how to be consistent with one’s own values in such a situation. If one is genuinely trying to do the right thing, the facts of one’s “internal world” make no difference to one’s intentions. One looks to the facts themselves.

Yet turning one’s perspective from the first- to the third-person, thinking not of the intention’s content but its structure as a mental state, the attempt to do the right thing appears very different. One may not think at all about one’s mental states and may try to examine only the actual facts of the matter. But from the third-person perspective, when a person sets out to recognize the fact of some subject matter, her efforts move toward generating a belief-state that accurately represents that fact, a belief-state that is true. She only has it in mind to determine the facts. She looks toward the world as though her looking were wholly transparent. But though the *content* of her intention may have nothing to do with her mental states, nonetheless, when described from the third-person perspective, she forms a series of such states in the process of pursuing what she does have in mind. She thinks *through* these mental states, not *about* them.

So when a person tries to recognize the facts her mental states will take on a certain pattern where she succeeds. When she recognizes forest green as the right color for the house, she generates the belief-state that forest green is the right color and the belief-state is true. When she recognizes she must pull the person from the wreckage, she generates the belief-state that she must pull the person from the wreckage and the belief-state is true. If one tries to do the right thing for its own sake, one tries to recognize the facts about what is right and respond with one’s actions as those facts require. When one succeeds, one will have belief-states that are true and actions and intentions that are consistent with those beliefs. Yet one’s actions and intentions will

be consistent with those beliefs because one tries to act on the basis of the facts as those beliefs represent them, not for the sake of consistency for its own sake.

So already it seems the intention to do the right thing has parts producing consistent beliefs and intentions when successful. Persons will be consistent where they succeed in recognizing the facts and responding to them in their actions and intentions. Thus when one acts in a way inconsistent with one's true beliefs, one possible explanation will always be that one is not really trying to do the right thing. First-order inconsistency will give defeasible evidence that one lacks that second-order intention.

However, explaining the "should" of internal consistency requires explaining why this relation holds even in cases where one is mistaken as to the right thing to do. Why "should" one live up to the dictates of one's religion even if one's religion is false? Completing this account requires further examining the content of the second-order intention.

4.2 Recognition and Best Judgment

Trying to see things as they are likewise appears not to be a simple mental state. It does not seem one has the intention to recognize the facts if, say, one views as accurate everything one is causally inclined or manipulated or forced into believing. Causal results do not track the truth. To see one's own views as merely causal results is thus to view them, pending further investigation, as inaccurate. So if one intends to determine the facts of some subject matter, one does not allow this to happen to one's reasoning. One tries to form one's views on the basis of the facts themselves.

Nor does one seem to intend to recognize the facts if one takes everything one carelessly or wishfully believes as accurate. One who intentionally reads a veridical science journal

uncarefully intentionally submits to having less accurate beliefs about, say, global warming than she could have with more or better effort. A person with the intention to have an accurate view, inasmuch as she retains that intention, cannot intentionally submit to that. Of course one may accidentally read off accurate information by carelessly reading a misprinted text. But one who commits to recognize the facts seems thereby already to commit to maximally overcoming her accidental circumstances and limitations and following the evidence with maximal accuracy as a matter of the constitutive structure of that intention. The commitment to recognizing the facts rules out intentional laxity because it is precisely its contrary mental state.

In attempting to determine the facts about some subject matter it seems one examines the facts bearing on that subject matter and follows their suggestion as to what is the case. Trying to determine whether the Earth is older than Venus is consulting some other facts in the world, whether an encyclopedia or the testimony of someone one trusts to tell the truth. Trying to determine whether the correct course of action would be to take the 5 or the 15 to get to Yosemite is to look at a map or ask a trusted person for advice. One does not allow one's work of determining the facts to be distorted or derailed by some force external to the facts themselves and one's reasoning about them, whether one's own psychology, others' manipulations, or the simple inertia of history. But one also does not allow that work to be undermined by one's own laziness, wishful thinking, or lack of rigor.

Yet though one may intend to follow the evidence, one cannot follow evidence perfectly merely by having the intention to do it. Causal interferences, especially one's own nature as an agent reasoning in space and time, make up fixed parts of the context within which one must determine what is the case. Intending to follow the evidence does not immediately free one from that context. So it seems clear that working to follow the evidence is already to try to block out

these interferences and track the evidence to the best of one's ability. If one sets out to determine, for instance, whether the evidence supports the reality of global warming, one does not genuinely have that intention if one does not do one's best to overcome whatever influence one's financial or political interests may have on one's evaluation of that evidence. If one does not try in earnest to do this or one is in denial about these influences, one is not genuinely trying to get it right, at all. An individual's ability to overcome such interferences and track the evidence will always be imperfect and some individuals have that ability to a greater extent than others. Nonetheless, one must try to the extent of one's ability to overcome these influences if one intends to recognize the facts themselves. The former attempt is a necessary part of the latter.

So to work to recognize the facts is to follow the evidence as closely as one is able. One does not think at all about one's "inner life" as one attempts, say, to determine whether the earth's average temperature is increasing due to the use of fossil fuels. One reads scientific journals, talks to experts or does the analysis oneself. Yet the extent to which one is directed away from one's mental states goes much farther than just declining to consult them. Trying to follow the evidence entails actively working to block out the causal influence of those states over one's thinking. One actively tries to overcome whatever influence one's financial or political interests may have over one's views about global warming and looks to the facts themselves. Part of trying to work out the facts is to work to maintain the transparency of one's beliefs where causal influences may compromise it.

But once again, the project of determining the facts has a different character from the third-person than from the first. From the first-person, one arrives at the correct view of the facts themselves by following the evidence as closely as possible. One reads climate science journals, interprets what they say to the best of one's ability, and works to come to an accurate view about

global warming by following what they say. Guided by evidence, one tries to recognize the facts themselves.

Yet one is not just a thinker but a person. One's attempt to come to a correct view of the facts is also the activity of an individual with mental states. As one reads a scientific journal and tries to understand and accept what it says, one generates a series of belief-states about global warming the content of which tracks that journal's testimony to a greater or lesser extent. This happens even though one does not think about one's inner life, at all. As one works as best one can to follow the evidence, one's efforts move toward forming a belief-state maximally consistent with that evidence given the scope and extent of one's abilities. If one works in earnest to view things correctly, one forms one's *best judgment*. One forms that best judgment not thinking of it *as* a judgment, as a mental state. One forms it without thinking of it at all. Rather, forming one's best judgment constitutes the mental background process through which one works toward the clearest view one can. From the first-person perspective, one works to the best of one's ability to view the facts as they are. The production of one's best judgment is incidental to the content of the intention and yet necessary to the intention's structure.

For instance, an investigator into global warming may be unable fully to understand the journals she reads or the lectures she hears or may have incorrigible beliefs forcing her to understand the evidence compatibly with those beliefs. As a result, working to the extent of her ability to follow the evidence may not produce in her exactly the same belief-states as would be produced in an expert working with the same material. Nonetheless, the relation each person's best judgment would bear to her intentions, her abilities, and the evidence she recognizes would remain the same. No less with the expert than with the novice, it would be the judgment that follows as closely as possible, given her abilities, the evidence she recognizes. Neither person

will think of it in the first instance *as* her best judgment. It will seem to be the way things are as far as she can make them out. From the third-person perspective, however, that view of things will still be her best judgment, even if that judgment is also true. This is no less the case if she takes up a third-person perspective on herself, assessing herself as one person with limitations.

So when one has the second-order intention one tries to recognize the facts about what is right and respond to those facts with one's actions. But in the course of trying to recognize those facts one forms one's best judgment and attempts to respond to its content. These are two aspects of the same intention, only seen from the first- and the third-person perspectives.

It is important to be clear about the kind of "best judgment" one produces in trying to determine what is right. One makes a "best judgment" in other senses when one judges about a question as though the question were not important or the evidence were too muddy for a definitive answer. When trying to work out the facts, one produces one's "best judgment" in the sense of the judgment that is the best one is able to produce, the way one judges concerning matters of the highest importance. One tries to the farthest extent of one's ability to follow the actual evidence. Unlike the previous two senses, a person judging in this way does not think about the extent of her abilities, at all. She thinks about the facts themselves and tries as best she can to see them clearly.

"Ability" should be clarified, as well. It is not that trying to do what is right is setting on a heroic quest for the right and the true. Two kinds of ability can determine the character of a person's best judgment. One's abilities to recognize evidence, understand what it suggests, identify instances of coercion, overcome causal interferences, and so on, are the primary limitations. But the time and resources available to a person within the larger context of her responsibilities constrain her, as well. It may not matter much within the big picture of the

things a person must do whether some particular action or belief is correct. While rushing her pregnant partner to the maternity ward, taking time to work out the best possible parking space would seem to most persons' best judgment entirely the wrong thing to do. Though the question may occur to her, a person will probably not arrive at the best judgment she should immediately re-examine all her political convictions while her boss is threatening her with unemployment and thus the inability to feed her family. One does not necessarily fail to act on one's best judgment when one makes judgments of priority. A person may judge with the utmost rigor available to her that not enough hangs on a particular question, that her chances of resolving it are not likely enough, or that the opportunity cost of pursuing it is too great to merit investigation, at least at a given time. She judges that her resources must be used elsewhere. She does this not at all contrary to her attempt to do the right thing but precisely *because* of it.

So working to overcome causal interferences and follow the evidence does not entail undertaking a heroic quest. It entails doing the best one can to do the right thing and see things the way they are.

4.3 Best Judgment and First-Order Consistency

Trying to do the right thing requires forming one's best judgment about what is right and trying to act on it. This structure appears to explain the relation between a person's first-order consistency and her possession or lack of the second-order intention and in turn her capacity or incapacity for right action. Right action must be done out of the intention to do the right thing. Having that second-order intention is forming one's best judgment and trying to act on its basis. One's judgment may be true or false. Yet a person necessarily will *attempt* to act on the judgment's content if the judgment is the best she can produce and she genuinely has the second-

order intention. She will make this attempt because her best judgment presents the facts themselves as far as she can make them out.

So if the religious racist comes to understand what human equality actually entails or that something she does or intends is racist and if she successfully reforms her racist intentions, her intentions necessarily *will* be consistent with her beliefs about what is right. But they will not be consistent because she necessarily cares about consistency, but because, from her perspective, she tried to do the right thing. The person try to do the right thing normally thinks not at all about the consistency or inconsistency of her mental states as such. Doing something merely because she *believed* she should do it would be precisely what she would try *not* to do. Instead, she thinks about the facts about what is right and whether she is living up to them. Nonetheless, her attempts to do the right thing result in the consistency of her attitudes when she succeeds, even by her own lights.⁴

5. First-Order Inconsistency as Evidence and Return to Cases

The structure of the second-order intention thus explains why a person's first-order inconsistency always gives reason to believe she lacks that intention. If successfully acting on one's best judgment necessarily produces first-order consistency, any instance of first-order inconsistency can have come about in one of two ways. Either one tried and failed to work out the facts to the best of one's ability and act consistently with them, or one did not try. Either one pursued the second-order intention and for whatever reason failed, or one lacked the second-

⁴ Kolodny himself has alluded to this state/content distinction as regards beliefs (Kolodny 520). Yet where he focuses on beliefs appearing to persons as facts, this argument focuses on the other direction the distinction runs. One has a belief where one has a view as to the facts, and so attempts to see the facts as they are result in the formation of beliefs as responsive to the evidence as one's abilities allow.

order intention. So first-order inconsistency gives defeasible evidence in every case of the lack of the second-order intention. That evidence may be defeated by other facts of the situation

This account seems to make sense of the cases of the religious racist and the good-natured egalitarian quite well. Several different explanations can be formulated for either person's inconsistency other than her unconcern with doing what is right. The good-natured egalitarian may simply be so good-natured she immediately does whatever the community thinks is right and thus does what is compatible with egalitarianism. The religious racist may genuinely lack understanding of the racist nature of the things she says or does. Both explanations are compatible with a person's trying to do the right thing. These factors stand in the way of her actually intending what is right by her own lights or intending what she thinks is right on the appropriate basis. One may still wish either individual were otherwise. But the assessment of such persons does not appear to be the same as if she were aware of her inconsistency, made no effort to reform, and no mitigating condition were present for why she did not make that effort. In the presence of those three conditions, it is evident she is inconsistent in the third way identified before, at least at that moment. She lacks the second-order intention to do the right thing. In many cases, one may be glad a person has not carried through on what she thought she should do. Yet it seems one would still find something amiss. Amoralism or arationalism explains that sense of something amiss. Where it genuinely obtains, it renders a person's actions ingenuine as right action, however right or wrong they may appear on the surface.

6. Summary of Theory

Right action, I have tried to suggest, must be done because it is right, which entails it must be done out of the intention to do what is right for its own sake. Trying to do what is right

for its own sake requires forming one's best judgment about what is right and trying to act on it. Where this intention succeeds, even by one's own lights, one's first-order intentions will be consistent with one's beliefs about what is right. One explanation available but defeasible in every instance of first-order inconsistency thus is that the person in question is not trying to do the right thing.

This paper set out to try to solve the question why persons are appropriately subject to criticism in every case wherein they fail to act or intend consistently with their beliefs about what is right, including cases where their beliefs are false. The solution I propose is that every instance of first-order inconsistency is defeasibly problematic because every instance of first-order inconsistency gives defeasible evidence of a person's lacking the second-order intention, that she is not trying to do what is right. Not to intend what is right, however, is to lack a necessary condition for all right action, as such. Declining to try to do what is right thus is problematic in every case, as it invalidates any merely outwardly good behavior. Instances of inconsistency are defeasibly problematic in every case not in themselves but because they give defeasible evidence of this deeper condition. Now I will try to show how this account meets all of Niko Kolodny's conditions for a successful theory of internal consistency.

7. Kolodny against Internal Consistency

The four conditions any successful theory must meet are not to involve bootstrapping, to be plausible, not to require persons to reason on the basis of the principles making up internal consistency, and yet to preserve the apparently genuine guiding role internal consistency plays in believing and intending. I will go through these in sequence.

7.1 Bootstrapping

John Broome was the first to call internal consistency requirements into question on the basis of their leading to “bootstrapping.”⁵ The problem is just this. Suppose one should intend what one believes one should do because it is right to intend or do what one believes one should do. Then, for any belief one has, one should intend to act on it. If this is supposed to explain the universality of the requirement, clearly the explanation fails. Kolodny repeats Broome’s criticism of the requirement and shows that even Broome’s own solution must run afoul of it.⁶

7.2 Plausibility

In Kolodny’s view, the reasons one could bring forward in favor of internal consistency are simply implausible as accounts of the universally problematic character of internal consistency. First, there is no general instrumental reason to be internally consistent. Intending what one believes one has most reason to do will not always put one in accord with the facts about what one has most reason to do. Nor could internal consistency be a good general policy for getting it right most of the time about what is right. If one could intend with perfect accuracy by sometimes ignoring one’s own judgment, why not ignore it? Moreover, even if consistency is a good policy in the long term, why be consistent in a particular case if one could do better on the whole by being inconsistent?

⁵ Beginning in John Broome, “Normative Requirements.” *Ratio* (1999) 12: 398-419.

⁶ In brief, “wide-scoping” the requirement, so that it can be satisfied not only by intending what one thinks is right but also by giving up the belief, gives license to clearly non-rational intention. Rational intention is undertaken in view of what is right. Giving up one’s belief that one has reason to do something just to make oneself consistent clearly is not rational action but rather a form of bad faith. (Kolodny 2005, p. 527)

Nor does it seem sufficient to suppose one commits to internal consistency just in virtue of being a believer and an intender, citing Davidson and Korsgaard as examples.⁷ One might think a rational agent as such cannot ask herself whether she should follow her beliefs about what is right without undermining her own nature.⁸ Continuing to follow these principles just constitutes rational self-preservation. Even if true, Kolodny says, one clearly does not cease to be a rational agent by one instance of *akrasia*. One might no longer count as a rational being if one continually failed in this way, but each failure does not entail rational death. Nor, Kolodny thinks, would it suffice to think that internal consistency just *does* have some overwhelming value persons must maintain and forward regardless of the consequences for their other reasons. This idea seems to him “precious and unreal.”⁹

Finally, one might suppose internal consistency to be a kind of “executive virtue.” Intending to act on one’s beliefs about what is right manifests a certain admirable character. Internal consistency thus would appear similar to the virtue of courage. However, there are many virtues, and though expressing their presence in one’s character may give one some reason to do something, none of them seems to always give decisive reason, as it would have to in order to explain the “should” of internal consistency. Kolodny concludes no substantive reason has been found, nor is any likely to be found, to make sense of that “should.”

⁷ Donald Davidson. “Incoherence and Irrationality” in *Problems of Rationality*. Oxford University Press, 2003: 189-98. ; Christine Korsgaard. *The Sources of Normativity*. Cambridge University Press, 1996.

⁵ Kolodny, p. 545.

⁹ Ibid.

7.3 Persons Cannot Act Rationally on the Basis of Internal Consistency

Third, Kolodny argues that internal consistency cannot have substantive reasons in its favor because persons would have to intend and act irrationally to fulfill the requirement. For instance, to move from “I believe I have reason to phi” to intending phi by way of the principle “It is right for me to intend or do what I believe I have reason to do” does not constitute rational action. In rational action, one undertakes an action because one has reason to do it. In this case, one would undertake the action just to bring about one’s internal consistency. Consistency as such cannot play the role in rational action that rightly belongs to the facts about one’s reasons, about what is right. Complying with the internal consistency requirement must not require one to think about or pursue internal consistency, at all.¹⁰

7.4 Apparently Genuinely Attitude-Guiding

And yet, Kolodny thinks, despite the failure of all these attempts, one cannot reduce the apparent first-person applicability of internal consistency to a mere third-personal evaluative attribution. Consistency seems to be genuinely intention-guiding in every case. An adequate theory must account for this guidance.

For instance, Scanlon at one point thought internal consistency could be the state of well-functioning for the cognitive and volitional apparatus.¹¹ Even if false, an intention following a person’s beliefs about what is right is still in a way good as an intention. Just as a knife is good

¹⁰ Kolodny’s made this argument with respect to believing in accordance with one’s beliefs about evidence. I have applied the same form of argument to inconsistency with one’s beliefs about what is right.

¹¹ T.M. Scanlon. “Metaphysics and Morals.” *Proceedings and Addresses of the American Philosophical Association* (November 2003) 77: 7-22.

as a knife when it can cut well, perhaps a volitional apparatus is good when its intentions are consistent with the cognitive apparatus' beliefs about what is right for the person to do.

Yet once again, this answer cannot ground intention-guidance in every case. Even if a well-functioning volitional apparatus is something good to have, this still is just one reason weighing against others, thus can be outweighed. Moreover, a knife's failure to cut well is not something the knife has done wrong. Things just turned out that way. But when, say, one refuses to intend the unpleasant requirements of one's beliefs, one fails in some way even if those beliefs are false. Internal inconsistency is a person's failing, not just a breakdown within a system.

Finally, theorists may attempt to argue that part of the constitutive purpose of beliefs and intentions is to be consistent with evidence- and reason-beliefs, respectively. But Kolodny says the constitutive purpose of beliefs clearly is to track the truth, which consistent beliefs and intentions can fail to do. That is the whole problem. So how could the functional purpose of truth-tracking make sense of the "should" of mere internal consistency? Still, he says, a compelling story remains to be told as to how the aim of the cognitive and volitional apparatus can be the facts while each particular belief or intention fulfills its function in mere consistency. One may still be able to argue that one must work to have consistent beliefs and intentions because they contribute to the whole apparatus' purposive organization toward the truth.¹² In a moment I will try to show that the account presented in this paper actually is this kind of account.

Kolodny concludes from the above arguments that no theory has managed to explain why internal inconsistency in one's attitudes is always problematic. Now I will try to show how the above account fulfills all these conditions.

¹² Kolodny, pp. 552-3.

8. Satisfying Kolodny's Conditions

8.1 Bootstrapping

Nothing has been said here to suggest either that internal consistency or the second-order intentions have value in their own right, nor that any particular consideration counts in their favor. It has only been suggested that if one has the second-order intentions and those second-order intentions or some of their component parts succeed, they result in internal consistency among one's first-order attitudes. If one has the second-order intentions, one intends to have attitudes that will be consistent if that intention is realized. But having this consistency-producing intention is just part of what it is to have the second-order intentions as a whole, not something it is right to do. They may not be things worth doing in their own right in any respect.

Yet even if the conjecture as to what makes a person's lack of the second-order intentions problematic turns out to be true, this account still would not be subject to bootstrapping objections. I have not suggested it is right to intend to do what is right because it is right, only that it seems a likely explanation for what it is to do any right action because it is right. If one does not do what is right unless one does it because it is right, it seems one who lacked the second-order intention would fail to do the right thing, regardless of the outward rectitude of her behaviors. One may be subject to criticism for lacking the intention necessary to any right action as such. But merely having the intention in and of itself would not suffice for right action.¹³

¹³ This structure can also respond to Broome's wide-scoping solution to the bootstrapping problem. Any agent with the second-order intentions intends for her belief to track the evidence and her actions to track what is right. Thus she intends to form true beliefs about what is evidence and what is right and to respond to those true beliefs with her other beliefs and her actions. She is not interested in having a set of wide-scope consistent attitudes that hangs together but perhaps swings clear of the facts. She wants the facts themselves. Nonetheless I have suggested she necessarily pursues this intention through forming her best judgment about what is evidence and what is right and trying to respond to that best judgment with her beliefs and actions. So though she may not care about it as such, her second-order intentions are not satisfied unless she is in an end-state in which her first-order attitudes have *narrow-scope* consistency. Where she lacks narrow-scope consistency, it thus is a sign that she may lack the second-order intentions. When one has the second-order intention, one intends to bring about *correct* attitudes. Yet in doing so, one thereby intends to bring about attitudes that are, from the third-person perspective, narrow-scope consistent.

8.2 Plausibility

I have not suggested internal consistency or the second-order intention are right to have in themselves, so this account avoids Kolodny's plausibility criticisms. He criticized the rational self-preservation and executive virtue views on the grounds that, even if those states have some reason or value in their favor, the reason or value probably does not overwhelm every competing consideration in every case. But I have not said consistency or trying to do the right thing have any reason or value in their favor, at all. I have suggested instead they are necessary but not sufficient conditions for any reason- or value-guided action, as such.

If the complete account I have outlined is correct, what makes the absence of internal consistency in a person problematic is the same old rightness of reasons that Kolodny favors. It is only that the conditions for doing the right thing are complex and include the second-order intention to right action. If that view is plausible, then my account of the importance of internal consistency would seem plausible, as well.

8.3 Persons Cannot Act Rationally on the Basis of Internal Consistency

This account in no way requires persons to reason or to act on the basis of internal consistency in order to fulfill its requirements. The principle, "I should intend what I believe I have most reason to do," need play no role in an agent's reasoning for her to follow them. In fact, a person who intends to do what is right will *not* move from "I believe φ is right" to intending φ . Rather, in the course of forming her best judgment, she will come to view φ as actually right and in the course of trying to do the right thing she will try to actually φ . She *will* have first-order consistent first-order attitudes if she succeeds and this consistency principle will describe the way her attitudes are consistent. Even so, her concern is not consistency but correctness.

So in fact this structure provides exactly the account Kolodny hypothesized as to how the parts of a system constitutively directed toward truth could perform their function when the system is merely consistent. They do so when from the perspective of the agent having formed her best judgment, thus from the perspective of her cognitive system working at peak capacity, her attitudes *are* correct, thus are performing their function. If she has formed the best judgment she was able to form, then, from her perspective, thus from her cognitive system's perspective, that judgment *is* true as far as she and as far as it can see. She works toward peak capacity as part of the second-order intention. Yet from her perspective as investigator it is merely incidental that the state she works toward entails her cognitive system's functional success. From her perspective, thus the system's perspective, she works toward the facts themselves.

So, for instance, probably vanishingly few Roman gladiators acted on their belief, "I should obey the emperor and fight the person in front of me," because they were following the principle, "One must act consistently with one's beliefs about what one should do." The majority acted on their belief they should obey the emperor because, as best they could tell, obeying the emperor really was the right thing to do. They did not, as it were, view things as we do and then tack on some insane beliefs about the authority of the emperor which, when combined with their commitment to their own consistency, required them to undertake drastically wrong actions. To the contrary, when they saw the emperor most of them saw a god whose will was, most of them thought, absolutely to be obeyed. They did what we see as wrong because they thought they otherwise would fail to do what was right. We can say they were mistaken to believe and do these things. But we cannot say all of them lacked the second-order intention.

8.4 Genuinely Intention-Guiding for Persons

This account seems to fulfill Kolodny's third requirement in the same way. An adequate account would explain why a person's beliefs about what is right must guide her intentions from the first-person perspective. Why must a person consult her beliefs to determine what to do?

This account asserts no special value or reason to explain why persons must guide their intentions in this way. One must guide one's intentions by one's best judgment not because it is right but because doing so is part of the attempt to do the right thing, which is part of any right action, at all. One must intend what one believes one should do not because it is one's own mental state but because, as far as one can tell, the action itself is right. One "must" intend in this way because that is what a person with the second-order intention necessarily *does*, as such. One tries to do what one's best judgment represents as right, however *that* one tries to act on one's best judgment is an incidental though necessary third-personal concomitant of one's first-personal intention to do what is right. One guides one's intentions by one's beliefs insofar as one guides one's intentions by what is right and true, so far as one can tell. The principles making up internal consistency guide one inasmuch as one intends what appears to one's best judgment right because it is, one believes, right. One need not represent those principles at all to follow them. One only has to attempt in earnest to do the right thing, whatever that may be.

9. Conclusion

It seems the question why one should have the second-order intention to bring consistency to one's first-order beliefs and intentions has no answer. Instead, one must have an intention that necessarily results in consistent attitudes where one succeeds if only by one's own lights. That intention is the intention to do what is right for its own sake. One must have it

because without it one cannot do anything that is right *because* it is right. But all right action must be done because it is right to count as right action.

This second-order intention necessarily results in consistent first-order attitudes because of what it is to pursue that intention. Trying to do the right thing is trying to determine what is right and do it. No part of the content of this intention mentions anything about one's mental states. If one genuinely tries to do the right thing, one has absolutely no interest in the contents or structure of one's mental states merely as such. Yet the content of one's intention is not the same as the structure of that intention mental state. Though one does not think *about* one's beliefs or intentions at all in trying to work out what is right and do it, one nonetheless pursues that attempt *through* a series of belief-states. As one does one's best to see the facts as they are, one's efforts move toward generating a belief-state that is as consistent with the evidence as one's abilities allow. That is one's best judgment. A person who tries to do the right thing must try to act on that best judgment not because it is her mental state but because that judgment, viewed from her perspective, presents things the way they in fact are, as far as she is able to make them out. Where she succeeds even only by her own lights, her beliefs and intentions will be consistent, not because she cared about consistency but because she sought to act rightly.

Instances of inconsistency thus can have many explanations, including a person's incomplete understanding of her beliefs or her lack of time for considering their implications. These explanations can defeat the problematic appearance inconsistency otherwise would give. Nonetheless one explanation always available unless defeated is that the person lacks the second-order intention, that she is not genuinely trying to do what is right.

This account seems to avoid all the pitfalls to which Kolodny shows the going theories to succumb because it posits no reason or value at all in favor of consistency. There is no problem

of bootstrapping or of giving a plausible reason because no reason or value counts in favor of consistency. There is no need for persons to reason out how to make their beliefs and intentions consistent because under this account persons should *avoid* doing precisely that, as that is not what one does if one genuinely tries to do what is right. Nor is there need to explain why a person's beliefs can and should guide her intentions. For the person genuinely trying to do the right thing, what guides her when she follows her best judgment about what is right is not her beliefs, as such, but what is right and what is true as far as she can make them out.

Consistent attitudes are not right attitudes, but right attitudes are consistent as well as right. One may have the former but lack the latter. Yet if one lacks the former, one necessarily lacks the latter, as well. But one must in every case intend to reach a consistent state because one must in every case try to do what is right.

CHAPTER 2

Right Action and the Scope of Judgment

1. Introduction

Why be consistent when you can be right? If you acted on your best judgment, you could fail to do the right thing or do the wrong thing believing it to be right. So how could consistency of this kind be required? In my view, one must always form and act on one's best judgment about what is right because doing so is a structural part of right action itself. Ordinary language and practice suggest that right action must be done *because* it is right, and doing something because it is right in turn seems to necessitate that one do it out of a second-order intention to do the right thing for its own sake. But *what one does* when one tries to do the right thing for its own sake, I argue, is form one's best judgment as to the right action and try to act on that best judgment. One does not have it in mind to act on one's best judgment when one sets out to do the right thing. Rather, forming one's best judgment and trying to act on it is the mental process *through which* one pursues what one does have it in mind to do. It is the third-person perspective on one's first-personal attempt to do the right thing for its own sake. Acting on one's best judgment, then, is a necessary though not sufficient condition of right action as such. This appears to follow just from the ordinary concept of doing the right thing.

Yet here the sufficiency objection arises. Even if this account follows from the ordinary concept of right action, is it adequate to *applications* of the concept? Are there not clear cases of persons doing the right thing even against their best judgments? Nomy Arpaly argues exactly this in her *Unprincipled Virtue*. She brings forth several cases that seem to be examples of persons doing what is right against their best judgments. She then claims we are forced to view these persons as failing to do the right thing unless we give up the necessity to right action of

acting on one's best judgment. Her most powerful example is the titular character from Mark Twain's *The Adventures of Huckleberry Finn*. Huck helps his formerly enslaved friend Jim escape to freedom even though Huck believes that to do so is to become accessory to theft. Any view that entailed that Huck fails to do the right thing here would be mistaken.

I argue this view lives off an unfounded assumption about the scope of judgment. Much of philosophy tends to class everything other than *deliberative* judgment as merely causal psychology, as "passion" in Hume's case or "drive" in Freud's. Yet this taxonomy of reason and psychology seems mistaken when we broaden our philosophical diet of cases. I agree with Arpaly that we frequently act on reasons without deliberation. But our non-deliberative, reason-responsive actions seem to exhibit the kind of complexity, articulation, and control we would not expect of the passive outcomes of causal mechanisms. Rather, these reason-responses appear to be intentional actions guided by a kind of immediate, non-deliberative judgment existing alongside the more familiar abstract, deliberative kind. We already recognize in purely theoretical reason a distinction between abstract and immediate judgments, between our convictions and what we see. There seems no reason at all to suppose that the same distinction cannot exist within the practical domain, as well.

As I will try to show, this distinction allows us to see three things. First, it helps us make sense of the kind of complex, reason-guided action we all engage in over the vast majority of our lives. Even the most reflective of us deliberate over what we should do in a relatively small number of situations. Without the distinction between abstract and immediate judgment, we would be forced to see ourselves as continually driven into these actions by causal mechanisms. On my view, we see the situation before us, we see what it requires, and we actively respond to that requirement. This picture seems more adequate to the phenomena. Second, distinguishing

abstract and immediate judgment lets us make sense of the conflict, familiar from everyday life, of seeing something as right in the moment that one abstractly believes to be wrong. Huck Finn seems to be the paradigmatic example of this conflict. Understanding Huck in this way seems a better analysis of his situation and his deep goodness than can be formulated if we only recognize deliberative judgment, and to be more faithful to Twain's actual text, as well. Third and finally, I will try to show that my account of right action, with this broadened conception of judgment, tracks our assessments of when a person doing something one colloquially calls "acting against her best judgment" counts as doing the right thing, as well as the reasons for those assessments. Thus my account of right action seems not to be a mere deductive construct from our ordinary concept of right action, but to be materially adequate to our everyday applications of that concept, as well.

2. Huck Finn as Putative Case of Right Action Against Best Judgment

Having been raised in antebellum Missouri, Huck was taught that slaves are human property belonging to another human being as fully and with the same rights of disposal as one has over owned objects. To aid a slave in escaping would be the same as aiding a thief in larceny. Huck believes these things. Yet he finds himself riding down the Mississippi with his enslaved friend Jim as both attempt to escape the circumstances of their lives. On the way, Huck finds he cannot bring himself to turn Jim in to the authorities even though he believes he should and that he commits a sin in failing to do so. There are two such episodes, though Arpaly elides them. In the first, Huck resolves to turn Jim in but is unable to go through with it after Jim calls Huck his best and only friend. In the second, Huck decides to send a letter to the woman from whom Jim has escaped telling her where to find him. Yet after thinking about how Jim had been

such a good friend to him, Huck exclaims, "All right, then, I'll GO to hell," resolves to give up trying to be good and continues with Jim down the river.¹⁴

As Arpaly construes these cases, Huck's best judgment, arrived at through deliberation, is that the right thing to do would be to turn Jim in. Yet Huck cannot bring himself to do such a thing to someone he sees as a person and as his friend. As Arpaly has it, Huck suffers weakness of will with regard to his best judgment. Yet his act is about as good as acts can be. Thus, she thinks, it must be possible to do the right thing even against one's best judgment. A person's actions can count as right even if they result from the causal overwhelming of her resolve by some part of her psychology. Arpaly thinks that to deny this is to say that Huck has not done the right thing in helping Jim escape. But no one should say this.¹⁵

Arpaly's argument seems very strong if we are not careful about the distinctions we employ. Yet when we look closely at the different things one can mean by "feeling," "response," "reaction," and "judgment," that strength seems to diminish.

3. Feeling, Response, Reaction, and the Scope of Judgment

Arpaly's analysis of Huck's case relies on an unexamined assumption she shares with much of the philosophical tradition about what kinds of mental activities count as judgment and which as mere psychology. She assumes that all judgment entails deliberation, and that everything other than deliberation is the passive outcome of "unconscious" psychological processes. As we will see, that assumption is deeply untrue to common experience. Indeed, we seem both to form and to act on *non*-deliberative, immediate judgment over by far the vast majority of our lives.

¹⁴ Mark Twain. *The Adventures of Huckleberry Finn*, Chapter 31.

¹⁵ Nomy Arpaly. *Unprincipled Virtue*. Oxford University Press, 2003, pp. 69-95.

Arpaly observes that deliberation occurs over an expanse of time. Yet there are clear cases of people acting for reasons without deliberation. Thus, she argues, judgment of any kind that p is a reason must be unnecessary for acting on p . It must be possible to act on reasons one does not on *any* level believe in through merely psychological, “unconscious” processes. Thus it must be possible to do the right thing without judging that action to be right.

Arpaly uses two examples as evidence. First, a tennis player wholly engaged in a match does not deliberate at all about what she should do. She responds immediately to situations as they arise, doing what they call for without stopping to cogitate about it. Yet despite the absence of deliberation such a person clearly acts for reasons. Second, Arpaly sets forth the example of Emily, a graduate student ill-fitted to her department and her discipline who still persists in them out of a sense of vocation to the work. One day, finally overwhelmed by frustrations caused by that ill fit, Emily rushes out of the department never to return. Yet it turns out that the life she builds afterward is the right one for her. Arpaly says that Emily acted for the reasons she had even though she acted out of emotion and against her best judgment.

Arpaly contends that Emily and the tennis player are the same kind of case. As she understands it, both individuals act on reasons “unconsciously,” which for her entails that they do so without judgment of any kind. Each is passively brought to her action, she thinks, through psychological processes that do not involve judgment. In Arpaly’s view, being brought to action by these processes suffices for doing the right thing because it is right, thus for right action.¹⁶

I think it is clear that these are not the same kind of case, at all, and that there is no reason to believe that everything non-deliberative must be merely psychological. Arpaly’s taxonomy of the conscious and the unconscious, judgment and psychology, appears to elide very different sorts of phenomena. This elision, which comes down to us largely unexamined from Hume and

¹⁶ Ibid, pp. 49-52.

even Freud, enables her to construe the causal overwhelming of a person's commitment to her deliberative judgment as the same kind of case as a person's non-deliberative employment of a skill. Yet this elision appears quite illicit. Arpaly's cases demonstrate it themselves.

It is true that both Emily and the tennis player act on "feel" or "feeling." Yet one often means radically different things by those words. One of those things does not require judgment, yet it seems clear the other does. The difference between these two kinds of "feel" or "feeling" lies in the way they affect the beliefs and intentions of the agent who has them, and, accordingly, in what senses one may "fail" to do as they direct one to do.

On the one hand are the "feelings" of anger, frustration, despair, pain, pleasure, and so on, with all the sub-taxonomies of emotion versus conation versus sensation. These kinds of "feeling" are all, at least in part, mere reactions to some state of affairs. Causal necessity is a necessary, if not sufficient, component of them all. The relationship between emotion, desire, and rationality is of course far more complicated than causally-minded thinkers like Hume and Freud have construed it. But to say that there is a causal element to emotion and desire is just to say that we all are at least partially *passive* before what feelings we feel and when we feel them, as well as before what courses of action they drive us toward. And this is quite obvious of sensation. We can tell ourselves not to feel them, but sometimes we still do, and they still urge us toward this or that action. This is the kind of "feeling" that overcomes Emily's resolve.

Now compare those feelings with the sense of necessity one gets when engaged wholly in an activity like building something, playing a sport, having a conversation with a friend, making art, playing music, or doing a logic or math problem. Nothing forces or impels one toward q from p and $p \rightarrow q$ or from 3 plus 2 to 5. Yet one *sees* that they must follow and, unless one succumbs to self-deception or some other form of bad faith about it, one does in fact draw the

right conclusion. Likewise, in construction, sport, conversation, and so on, one very easily could go wrong at any time. One could make the wrong move or just not move at all. Indeed, if one stops to deliberate or notices one is *not* deliberating, one often is much more likely to go wrong. Yet failing materialist naturalism, nothing *forces* one to make the right move. It is true that one “feels” the necessity of the right tool to use, the right move to make, the next word to say. But it is not the necessity of something’s exerting force on one toward some series of events. One has to respond to the necessity, oneself.

Thus the ways in which one can “fail” to do what these two kinds of feeling necessitate are profoundly different. With an emotion, a desire, or a sensation, one can only “fail” to do what the feeling necessitates in the sense that one does not do what that feeling would, in an uninterrupted course of things, impel one to do. One “fails” to do what one’s desire to eat necessitates one to do when one does not eat. Yet one will not go in the direction the second kind of feeling necessitates unless one actively determines oneself to go that way. One fails to draw the conclusion that 3 plus 2 necessitates if one draws 6 or does not draw 5. One fails to perfect the sculpture if one stops working too early or too late. But nothing *forces* one’s mind or one’s hand in these ways. Because of this, one can “fail” to follow the second kind of feeling in the sense of making a *mistake*. But this makes no sense with reference to the first kind. The first kind of “feeling” *impels*. The second kind *guides*.

Telling these kinds of feeling apart is difficult when thinking about one’s own motivations or even another person’s. Self-deception is probably more readily available here than anywhere. Nonetheless, they *are* different, and yet Arpaly conflates them. She confuses active *response* to a situation with passive *reaction* to some element in it.

Response seems clearly to necessitate a non-deliberative form of judgment, where reaction does not. One sees immediately what is to be done and one does it. Through one's awareness of the demands of the situation, one senses that φ is the right way to go and one guides one's intention by that sense. One does not have to think about it at all, and yet the response is highly articulate and tightly controlled in a way we do not think of merely psychological, partially causal responses as being. This is especially the case for a skilled tennis player, conversationalist, sculptor, or mathematician. One's psychology will come to one's aid if one is well-conditioned to this particular task. But it does not on any level *force* one into action as events causing a reaction would. Thus we can distinguish the immediate, non-deliberative kind of judgment made by a tennis player from the abstract kind of judgment we employ and attempt to reach in deliberation.

Over the rest of this chapter, I want to show that this distinction allows us to make sense of three things crucial to the understanding of right action. First, we better understand everyday non-deliberative rational action. Second, we make sense of a familiar kind of conflict we experience between what we believe must be right and what we see as right in the moment, which makes no sense if we recognize only deliberative judgment. Third, we understand why we approve of cases like Huck Finn's in a way we do not when persons perform correct actions while thinking them wrong, yet do so out of a compulsion toward those actions or out of an interest in them other than their correctness.

4. Immediate Judgment and Everyday Non-Deliberative Action

I want to suggest now that recognizing immediate judgment about right action allows us to avoid certain problems having to do with everyday non-deliberative rational action that arise

for systems that do not recognize it. First, we can understand non-deliberative right action in a way that does not require us to count as right action things that seem clearly not to be. Second, the non-deliberative rational action we engage in over the vast majority of our lives need not be explained as the passive outcome of psychological mechanisms. Even if we do not deliberate when we do the right thing, we are not thereby *forced* to do it.

Say that Emily acted as she did about her graduate program with regard to some other issue. Say she just threw money at homeless people because she could not stand the sight of their suffering despite thinking it would be right to donate it to the local homeless shelter. But say that the shelter is corrupt and the money would not be used to house people, anyway. It seems one would be even less likely to think Emily had acted rightly in this case than in leaving her program. She did do the factually correct action, and she also was *caused* to do so by the fact that was the reason to do it, namely the suffering of a human being. Nonetheless, she did not do it *for* that reason. She was brought passively to that action by the circumstances of the situation. She did not actively determine herself to the action on the basis of its being the right thing to do.

By contrast, say one encountered a person who did not think at all about the larger social structures of homelessness, but who immediately gave help in this situation just because a human being asked her. She was in the situation, heard the person's request, and responded without a second thought. It seems one would regard this person as having done the right thing because it was right. In this case one would be faced with the moral equivalent of a tennis player, someone who does what is right without having to deliberate over what to do.

We encounter such persons constantly in the course of our everyday lives. They are, on the one hand, the morally decent people who surround us. They are also, on the other hand, each one of us, ourselves. Even the most reflective of us spend the vast majority of our lives

responding non-deliberatively to circumstances as they come along without stopping to think about them. We do what seems best in the moment. For the most part it is only when problems arise, when unsure of what to do, that we deliberate. We move around the world responding to the aspects of it that seem salient, completely certain at nearly every point of what we should do. Yet like those of a tennis player in a match, our actions are extraordinarily complex and controlled. So once again, to explain them in terms of psychological mechanisms seems a strained reading of the data. Outside of larger views about nature as a whole, there seems no reason to believe we are being driven by unconscious psychological forces in every case of rational action we do not undertake through deliberation.

5. Conflict Between Abstract and Immediate Judgments about What is Right

Second, the distinction between abstract and immediate judgment makes sense of a familiar kind of rational conflict as to what is right to do. One can encounter a conflict between two contradictory views of how things are, between the content of one's abstract judgment and the content of one's immediate judgment. Faced with this conflict, one can be *in* conflict oneself as to how to resolve it, can struggle over whether to maintain one's convictions or to accept the way things appear. We already view conflicts between abstract and immediate judgment as possible within the purely theoretical domain. I see no reason to think conflicts of this kind cannot exist in the practical domain, as well. Huck Finn, in fact, appears to give us the paradigm example of just such a conflict.

Consider the first person to discover a platypus. We can imagine she believed the old pieties about mammals, thus still believed in the abstract that no mammal lays eggs. Yet there was the platypus. On the one hand, it could not be. On the other hand, there it was. Such a

conflict is not merely a causal one between a person's reason, will, and psychology. It is a conflict of content between her abstract and immediate judgments. To resolve it, she would have to form her best judgment about which view to trust or as to how they may be reconciled.

Yet the need to resolve such a conflict does not follow simply from its existence. We can imagine someone encountering a conflict between her abstract convictions and what she sees but being indifferent to that conflict and thus not bothering to try to resolve it. It seems that if she did not care about the conflict she very well could go about her life with contradictory beliefs. We talk about people doing this all the time. So it seems a conflict of content of this kind would cause a person to struggle over it only if some other condition were met. One such condition would be that she happens to value her internal consistency for its own sake. Yet far more likely, she would struggle to resolve the contradiction if she were trying to see things as they are, if she had the standing second-order intention to have true beliefs. A person's struggle to reconcile a conflict of this kind thus would give defeasible evidence that she was trying to work out the facts. Failing to struggle in this way, on the other hand, would give indefeasible evidence that she *lacked* that intention, that she was not trying to see things as they are.

We now have the philosophical machinery to make sense of conflicts of this kind within a practical context. One can abstractly believe something to be wrong and yet immediately sense that it is right. Huck seems to be the ideal example of just this kind of conflict.

On the one hand, Huck believes it is wrong to help Jim escape from enslavement. He believes this on the testimony of every authority figure of the society he has lived in all his life. But on the other hand, Huck sees for himself both Jim's humanity and that they are friends. Yet friends help friends escape slavery. So Huck's learned conviction and his immediate sense of right and wrong conflict. Now, because he trusts the adults around him, Huck believes abstractly

that this “feeling” must be mistaken. If it were not, the adults would be *wrong*, and so far Huck seems unable to question whether the upstanding citizens of his society are right about the justice of slavery. Yet at the same time, Huck has an inchoate trust in himself and so does not look away from his responsibility to Jim. So Huck does not simply experience a causal conflict between his reason, will, and psychology as though as a spectator. He encounters a conflict of content between his abstract conviction and his immediate sense of what is right.

Yet Huck clearly does not indifferently pass over this conflict. He struggles intensely to reconcile these two testimonies, to form his best judgment as to what to do. He very easily could forego this struggle and turn Jim in despite seeing Jim’s humanity and their friendship. Huck could do this, for instance, if he cared about something more than doing the right thing, say, about remaining an upstanding citizen within his society. Or he might simply recoil psychologically from seeing Jim as a human being in fear of confronting his society’s moral sickness. Most of the adults in his society clearly have done one or both of these things. Yet Huck does neither. He sees Jim’s humanity and friendship and does not look away for the sake of saving social face, of personal comfort, or even what his society has told him is right. Huck struggles over what he should do despite his abstract conviction as to the answer precisely because of the strength of his intention to do the right thing. That is what kills his resolve and disallows him from looking away from Jim’s humanity and their friendship.

Both cases of Huck turning aside from his resolution to turn Jim in bear out this reading. In the first, Huck actually initially feels quite negatively toward Jim, viewing him as morally corrupt on account of Jim’s plan to buy or steal his family out of enslavement once he reaches freedom. Huck even repeats some extremely racist epigrams to himself. So he is at one with himself when he decides to turn Jim in, even singing as he sets out. But as he does so, Jim calls

out that he will never forget Huck for helping him, that Huck is both his best and his only remaining friend and the only white man who ever kept a promise to him. Huck says that Jim's words "seemed to kind of take the tuck all out of me," and that he felt sick. As a result of this loss of resolution, Huck finds himself unable to tell the authorities the skin color of his companion on the raft.¹⁷ Yet what undermines Huck's resolve clearly is not Jim's simply re-animating Huck's positive feelings toward him. The emotion that unsettles Huck is not fondness or anticipated pain, but *guilt*. Unlike some other emotions, one does not feel guilt unless one believes one should have done something, one cares about doing it, and one believes oneself to have failed to do it. So what undermines Huck's resolve is not the causal power of his emotions, but his awareness that he owes Jim his help and protection. He finds himself caught between his immediate view, his abstract conviction, and his own intention to do the right thing.

The second case confirms this account definitively. Huck writes a letter to the woman from whom Jim has escaped telling her where Jim is to be found. But before he goes to send it, Huck starts to think about his friendship with Jim and becomes conflicted again. Now he consciously resolves to settle this conflict once and for all. He begins *deliberating* about his obligations to a friend as opposed to doing what he has been told is right. He says he

got to thinking over our trip down the river; and I see Jim before me all the time: in the day and in the night-time, sometimes moonlight, sometimes storms, and we a-floating along, talking and singing and laughing. But somehow I couldn't seem to strike no places to harden me against him, but only the other kind. I'd see him standing my watch on top of his'n, 'stead of calling me, so I could go on sleeping; and see him how glad he was when I come back out of the fog; and when I come to him again in the swamp, up there where the feud was; and such-like times; and would always call me honey, and pet me and do everything he could think of for me, and how good he always was; and at last I struck the time I saved him by telling the men we had small-pox aboard, and he was so grateful, and said I was the best friend old Jim ever had in the world, and the ONLY one he's got now; and then I happened to look around and see that paper.

¹⁷ Twain, Chapter 16.

It was a close place. I took it up, and held it in my hand. I was a-trembling, because I'd got to decide, forever, betwixt two things, and I knowed it. I studied a minute, sort of holding my breath, and then says to myself:

'All right, then, I'll GO to hell' -- and tore it up.¹⁸

So, much in contrast to Arpaly's construal of Huck as overcome by his feelings, Huck *decides* in full awareness to disobey what he believes on the adults' testimony to be right, and he does so precisely on the basis of the fact that Jim is his friend. He cannot harden himself against Jim because he cannot find a way not to see Jim as a human being to whom he owes the duties of friendship, not because he simply cannot make himself stop feeling affection for him. He cannot resolve the one belief into the other, as they prescribe and proscribe exactly the same action. Yet he has to decide. The way Huck finally resolves the conflict is the way he has always resolved the conflict between what the adults say is right and what he himself sees fit to do. He trusts both the adults and his inchoate sense, and simply deliberately views himself as wicked rather than give up either testimony. Faced with a conflict he is unable to resolve by any other means, Huck does what any of us would do. He goes with the one that feels right. For Huck, that is to face any consequence rather than let down a friend.

So Huck *does* act on a judgment about what is right, namely the testimony of his eyes that Jim is not property but a human being and his friend. He is "unable" to turn Jim in because he intends to do what is right and thus refuses to look away from what he sees. Huck also acts on his *best* judgment in both cases. The question whether one has acted on one's best judgment is the question of the process one has gone through in arriving at that judgment, not *which* judgment it is. One's best judgment very well may end in irresolution. Yet the nature of one's best judgment is that one cannot resolve it by further thinking, either because one's own

¹⁸ Ibid, Chapter 31.

capacities are insufficient or because a decision is made necessary by the situation. So in certain cases, intending or even believing on the basis of one's best judgment actually may require one to make an arbitrary or emotional choice. Huck has reasoned in the best faith about what he should do, and he has tried as best he could to do it. If anything is acting on best judgment, it would be this.

Yet one may wonder why Huck thinks of *himself* as being overcome by his feelings, as weak and unmanly and even wicked in his inability to carry through his intention to turn Jim in. This *would* seem to suggest Arpaly's reading. I suggest that Huck misunderstands himself in much the same way as Arpaly misunderstands him. This is part of the pathos of the book. We do not merely admire the goodness of Huck's heart. We also feel a certain pity for him. I suggest that we feel that pity for two different reasons. First, Huck lives within a perverse ideology that twists his good heart against itself. He does not yet know that social consensus can be wrong, that adults can be in bad faith, and that a person can go against social consensus for the sake of doing the right thing and succeed. Yet he also trusts himself and refuses to look away from Jim's friendship and humanity. The only way he can reconcile this conflict is to view himself as wicked for following his inchoate sense of right and wrong as he continues good-naturedly doing "wicked" things. The second part is this. Huck recognizes that some part of him other than his abstract conviction keeps him from doing what the adults say is right. But because he still does not question their testimony, he views that part as pulling him away from what *in fact* is right, thus as unresponsive and even inimical to the truth. So Huck has to misunderstand himself. He has to understand the strength of his own sense of responsibility for Jim and his own intention to do the right thing as *weakness*. Otherwise the adults would be wrong. But so far he cannot accept that. So not only does Huck think himself wicked. He also

is forced to misunderstand his own moral strength as weakness by his trust in the moral strength of the adults around him.

Where for Huck anything that pulls him away from what the adults say is right must be weakness, for Arpaly anything that pulls us away from our deliberative judgment must be mere psychology. Where something conflicts with our abstract conviction, this can only be a *causal* conflict, not one of content. This picture causes her to read Huck the same way Huck does, as lacking strength of resolution. But this is untrue of mental kinds and of Huck. What sometimes looks like being overcome by a feeling is actually being unable to look away from the truth, in Huck's case the humanity of his friend Jim. Yet the only thing that can cause *this* inability is the power of a person's intention to do the right thing, which entails forming one's best judgment about what is right. The forms of judgment are just more various than many have assumed.

Now we have seen two things. First, one can act on one's best judgment without deliberation. Second, one can act on one's best judgment even if one abstractly believes oneself to fail to do the right thing. Now I want to attempt to show that this is not just a way to save the view that right action necessitates action on best judgment. Besides her assumption that judgment entails deliberation, Arpaly's argument also relies on our ability to refer colloquially to a variety of different cases, including both Emily's and Huck's, as a person "doing the right thing against her best judgment." Yet it seems one would not count *all* such cases as genuine right action. I now want to try to show that my account of right action, with the scope of judgment clarified as above, tracks the judgments it seems one would make in such cases, as well as the reasons for those judgments. My account thus seems not merely a deductive construct from the ordinary concept of right action. It seems adequate to ordinary applications of that concept, as well.

6. The Broader Account of Right Action Tracks Applications of the Concept

Early in her argument, Arpaly brings forth the example of Sam, a student who has not kept up with his studies over the semester. Sam deliberately concludes that he should do nothing social and only study until exams are over. Yet he cannot keep to this plan. He goes on being social, studying some amount, and finishes the semester with decent grades. Yet contrary to his judgment of the situation, Sam's grades would have been worse if he had gone through with his plan to be temporarily monastic. He also would have missed out on forming some social ties that will be important in his subsequent life. Arpaly classes this as an instance of doing the right thing against one's best judgment and as genuine right action, as such.

I suggest instead one does not yet know what to say about Sam, that his case is underdescribed. His story could be filled out in a way one would be inclined to call genuine right action. Yet it could also be filled out in ways one would not.

Here is a Huck-like story about Sam. Sam₁ has been taught all his life that the only thing that matters is personal material success, thus that the point of college is to make grades and future business connections. Deep down, Sam₁ has a sense for intrinsic values much wider than personal material success. He suppresses this sense, however, believing it misguided. Nor has he ever had the chance to employ that sense, as he has spent his entire life surrounded by people living by the values he has been taught. Yet say that in college Sam₁ discovers people who genuinely care about things and people for their own sake. He still believes abstractly that these people are wasting time studying history or philosophy or in social interactions not covertly about networking. Yet he still feels drawn to their different approach to things. Say that before Sam₁ turns off his phone to become a hermit they call to invite him to a party. He declines. Yet he soon finds himself conflicted over what to do. The sense of intrinsic values he suppressed all

his life will not leave him alone. He sees their way of life as better than the one he has been taught, though he still does not give up his deliberative conviction that it cannot be. Eventually, though he abstractly believes he should not, he puts down his books and heads over to the party. But say this is the best thing he could have done. He makes his first genuine friendships and does better on his finals than had he gone through with his plan.

It seems one would say that Sam₁ does the right thing in this case. Meeting these people was for him like the discovery of the platypus. He abstractly held to a certain ideology. Nonetheless, he discovered the possibility of a better life and did not look away from it so as to save face within his previous society or out of a psychological recoil from that possibility. One would think him wrong if he had looked away from or had not acted on that discovery just as one would think Huck wrong if he had looked away from Jim's humanity or had seen it and still turned Jim in, anyway.

Now consider two variations on the case. First, say that even though he was taught that social interactions only matter as a means of forming business connections, Sam₂ has a weakness for parties. On receiving the invitation, Sam₂ declines and tries to study. Yet he cannot stop thinking about the party. Finally, after a short struggle, he throws up his hands, heads out and never thinks about it again. Yet he makes better grades and social connections than he would have otherwise.

Second, say that Sam₃ is constituted just like Sam₁ in most respects. He has all the same experiences, makes all the same discoveries, and thus confronts the same conflict between his abstract conviction and his immediate sense as Sam₁. Yet say that Sam₃ differs from Sam₁ in that even though he has a sense of intrinsic values wider than personal material success, deep down Sam₃ cares more about living a rich and interesting life than about doing the right thing.

Since he has lived exclusively around people pursuing only personal material success, Sam₃ has never had the chance to live according to his wider sense of value. But now, having met people with a similar sense, Sam₃ leaps at the chance finally to live a rich and interesting life. He still believes he should stay in and work, and going to the party does appear to him as the right thing to do. Yet he completely disregards the conflict between his conviction and his immediate sense. He heads over to the party immediately, without deliberating about it is right, at all. He goes because living according to a wider set of values is more interesting than living according to personal material success alone, not because living according to those values is right. So just like Sam₁ he encounters the conflict of content between his immediate and abstract judgments. Yet unlike Sam₁ he does not struggle over it.

I suggest that in neither of these cases would one want to say that Sam has acted rightly. We see why, it seems, when we think of how these actions appear from Sam's own perspective.

From Sam₂'s perspective, because he has genuinely formed his best judgment, the correctness of being a hermit does not appear to him as merely a mental state that may be true or false. It appears to him instead as what he really should do as far as he can make it out. So when he fails to live up to that best judgment, he does not view himself as merely failing to be internally consistent. He views himself as failing to do the right thing. When Sam₂ goes to the party instead of working, he thus cannot view that action as correct, since his best judgment is that something else is the right thing to do. So it seems that even though going to the party actually is correct, Sam₂ cannot actively undertake that action *because* it is correct. He cannot go to the party *on account* of its being the right thing to do. He commits his efforts, his contribution to his action, toward doing what he wants, not toward doing the right thing as such. The

correctness of his action thus is not his doing. The facts about what is right merely happened to coincide with what he wanted to do. He made no effort to ensure their correctness.

It seems one would not want to say that Sam₂ acted rightly in such a case, even though he performed the correct behavior. One would say he did the right thing because he wanted to do it, not *because* it is right. He could not do the right thing because it is right because he did not view it as the right thing to do. The outcome of his best judgment, the fact about what was right, so far as he could make it out, was that it was not right, at all.

The conflict between the contents of Sam₃'s judgments does not appear from his perspective as a mere fact about his mental states, either. Instead, there appears before him a state of affairs he believes cannot exist. Sam still is of the conviction it cannot be right to go to this party. Yet it still directly appears to him to be right. He views the world as including and not including the same fact. But this cannot be how things are. So if Sam₃ were trying to determine the facts about what he should do, he would do his best to resolve this conflict. He would form his best judgment as to what is right. This would require him to struggle to resolve the conflict either by discarding one judgment or by finding some way to reconcile them. Yet Sam₃ does not concern himself with the conflict at all. He does not arrive at the fact of the matter as far as he can make it out. He goes to the party not because, from his perspective, it is among the things right to do. He goes to the party because, from his perspective, it will help make his life more interesting. He goes on account of its interestingness, not its correctness.

So though his going to the party may be correct, it seems one would not want to say Sam₃ acts rightly in going to the party, either. The correct behavior may have occurred, but *he* did not do the right thing *as such*, himself. He made no effort to make his actions correct, which would have required him to form his best judgment about what is right and act on it. The facts about

what is right merely happened to coincide with Sam₃'s best judgment about what was interesting. It thus seems Sam₃ cannot be credited with having acted rightly. Because he lacked the intention to do what is right, which would have required his forming his best judgment and trying to act on it, the credit for his performing the factually correct action belongs wholly to the world and not at all to Sam₃ himself.

I have suggested right action must be the outcome of the second-order intention to right action for its own sake and that one acts out of the second-order intention to right action for its own sake only if one forms and acts on one's best judgment of what is right. But this third clause has two parts. One must form the best judgment one is able to produce about what is right *and* one must act on it. Huck, as I have tried to show, does both of these things. Sam₁ does both, as well. Yet if one fails in either of these parts of the second-order intention, one does *not* do the right thing, even if one does the action that happens to be correct. That correct action may happen *through* one, but one does not actively do it oneself. Thus failing to form or to act on one's best judgment gives deductive evidence one has not acted rightly.

In previous stages of my defense of the necessity of internal consistency to right action, I argued that the structure of right action follows from right action's ordinary concept. Yet as I have tried to show here, this analysis tracks our ordinary *applications* of that concept, including the reasons for that application, as well. It appears not to be a mere deductive construct, but rather as materially adequate to what we say and do.

7. Conclusion

A certain reading of the example of Huck Finn seems to show acting on one's best judgment cannot be necessary to right action. Under Arpaly's analysis of the case, Huck is

causally overcome by his feelings when he helps his friend Jim escape enslavement as against his deliberative belief that he should turn Jim in to the authorities. Huck's emotions forcibly overwhelm his resolve to do what he thinks is right. Yet he still acts rightly. Right action, then, must not require action on best judgment. Views like mine seem to her to force us into the deeply problematic view that Huck has not done the right thing.

I have tried to demonstrate that this argument relies on an overly narrow view of judgment that Arpaly shares with much of the philosophical tradition. According to that view, all judgment about right action entails deliberation and thus any action done apart from deliberation cannot be done on the basis of the judgment that it is right. Yet this seems unfounded on the basis of Arpaly's own examples. A tennis player returning a serve while wholly immersed in her match and a graduate student overwhelmed by her frustrations running away from her ill-fitting program do not appear to be the same kind of case, at all. The former actively *responds* to her circumstances as she encounters them. The latter passively *reacts* to her circumstances through the emotions they cause in her. Active response seems to require that one immediately recognize what the circumstances require one to do. We seem to "see" or "sense" or "feel" such requirements all the time, especially when immersively engaged in complex activities like conversation or sport or mathematics. By contrast, mere causal reactions do not require recognition of any kind. So while it is possible in principle to explain active responses like the tennis player's in terms of causal mechanisms, it seems we have no reason to do so. It seems that we instead should recognize an immediate, non-deliberative form of judgment of what is right alongside the more familiar abstract, deliberative form.

Recognizing the existence of immediate judgment about right action gives us three results. First, we understand everyday rational action more adequately than if we assumed all

such action to be the outcome of psychological mechanisms. All of us deliberate over a very small subset of the rational actions we engage in, including moral actions. Yet these non-deliberative rational actions are highly complex in the way a tennis player's are. Explaining these in terms of causal mechanisms would likewise seem untrue to the phenomena. To the contrary, it seems we employ immediate, non-deliberative judgments about what is right and respond to the contents of those judgments without stopping to cogitate about what we should do. We do this constantly over the vast majority of our lives.

Second, immediate judgment allows us to make sense of conflicts of content between our abstract judgments about what is right and what seems right to us in the moment. Such conflicts are exceedingly common in theoretical contexts. There seems no reason to think they cannot exist within the practical domain, as well. Indeed, Huck Finn seems their paradigmatic example. Huck believes abstractly in the justice of slavery on the testimony of the adults of his society. Yet he also sees for himself that Jim is a human being and his friend and that friends should help friends escape slavery. Precisely because he is trying to do the right thing, Huck struggles to resolve this conflict between his abstract conviction and the evidence of his eyes. He struggles to form his best judgment of what he should do. In the end, he resolves the conflict neither by giving up his trust in the adults nor by giving up his trust in himself. He resolves it by giving up trying to be "good" by his society's standards and helping Jim escape to freedom. This seems an understanding of Huck both better in its own right and more faithful to Twain's text, yet it makes no sense if all judgment is deliberative.

Finally, allowing for non-deliberative judgment, thus for this form of conflict, suggests the accuracy of the broader account of right action as requiring action on best judgment. Where a person does something called "doing the right thing against her best judgment" in the way that

Huck does, as the resolution of a conflict of content between her judgments about what is right, it seems she acts rightly. Yet if her conflict is merely a causal one like those Arpaly describes, or if she faces a conflict of content and yet declines to try to resolve it, it seems she does not. We understand this difference when we approach these conflicts from the person's own perspective. From her perspective, a person's best judgment is not merely a mental state. It is the fact of the matter as far as her abilities allow her to see it. Thus, when a person acts contrary to her best judgment, she cannot view herself as doing the right thing. Since she cannot see what she is doing as right, she cannot commit her efforts toward doing that action *on the basis of* its correctness. Thus she cannot count as doing the right thing, herself, even if good events occur through her behaviors. Likewise, from her own perspective, a conflict of content within a person's judgments is not a mere characteristic of her belief-set, but rather the appearance of a state of affairs she abstractly believes cannot exist. If she sought an accurate view of what is right, that person would work as best she could to resolve this conflict, would form her best judgment. But from her own perspective, to decline to form her best judgment would be to decline to come to that accurate view. Whatever action she then performs, she thus cannot view it as correct, and so cannot commit to doing it *on the basis of* its correctness. Though she puts forth efforts toward doing the action, its correctness is accidental to her doing so. Thus she cannot count as doing the right thing, herself.

The above account of the structure of right action would predict exactly these conclusions and for these reasons. I argued first that right action must be done for its own sake and thus must be done out of a second-order intention to do the right thing for its own sake. Second, I argued acting out of that second-order intention is forming and acting on one's best judgment of what is right. As a result of this structure, a person's failure to form or to act on her best judgment gives

deductive evidence she has not done what is right. Yet the comparison of Huck Finn-like cases to cases of persons performing correct behaviors without forming or acting on their best judgments shows this, as well. So my account appears not merely to follow from an analysis of our ordinary concept of right action but to be adequate to our ordinary applications of that concept, as well. So if correctness alone matters, that would not undermine the requirement of action on best judgment but provide its foundation.

CHAPTER 3

Two Accounts of the Instrumental Principle

1. Introduction

Why be consistent if you are wrong? Questions about the Instrumental Principle (“If you intend end E and believe M is a means necessary to E, you must intend M”) are questions as to why one can ask a person “Shouldn’t you be M-ing if you’re planning to E?” where M is a means she believes necessary to E, even if one thinks it is not right to E. Following the principle in one’s actions and intentions often could lead one into incorrect action. So what sense, if any, can this “should” have? And why does the principle apply in *every* case?

The view that one has reason to be instrumentally consistent has evident problems. It does not seem to become correct to undertake an action merely because one believes that action necessary to an end one happens to have, whether than end is right or wrong. So what, then, could be the ground of the principle?

I do not want to suggest either that persons have or that they do not have reason to intend consistently. For all I argue here, a view of either kind may be true. I want instead to try to resolve the problem by altering the direction of inquiry. Rather than suggest persons have or do not have particular reasons to be instrumentally consistent, I argue instrumental consistency is in two different ways constitutive of acting for reasons, of right action, as such.

First, I argue, a person who does the right thing views her action as right and performs it because it is, she believes, right. In the process, she necessarily intends the means she believes necessary to her right action. One such action would be achieving her right end. Without taking the means necessary to an end, I suggest, a person cannot achieve the end, even if her actions causally produce a state of affairs resembling that end. Without instrumental consistency *she*

will not achieve the end through her own intentions, at all. Yet one who does not achieve an end, at all, cannot achieve the end because it is right, either. So persons carrying out right action *will* intend consistently with the Instrumental Principle as a matter of right action's constitutive structure. Yet they will do it not to be consistent for its own sake but to do the right thing. Particular instances of instrumental inconsistency thus give defeasible evidence a person cannot or will not do the right thing. If one cannot or will not do what achieving one's incorrect ends requires, perhaps one cannot or will not do what achieving right ends requires, either. A person reciting the Instrumental Principle in this sense does not inform one of what is right in a particular case. She reminds one how one must be to do what is right. One must be disposed to actually do it, to achieve it in *reality*.

The other sense is both epistemic and volitional. Within human social life, one presumes persons' actions to be appropriately subject to questioning as to their justification. This presumption would seem to make no sense without the further presumption that persons normally undertake actions viewing those actions as right. I will argue that within these presumptions one must also expect persons to view the means they believe necessary to their ends as right and intend those means. One must expect them to intend the means they believe necessary to their ends insofar as actual right ends necessarily count in favor of the means actually necessary to bringing them about. For persons to fail to do take the means they believe necessary would give evidence either of their unwillingness to accept the full implications of their own beliefs or of their unwillingness to act on what appears to them as right. Either way, their omissions would give evidence of their lacking the second-order intention to right action. But the second-order intention partly constitutes all right action as such. One "must" be

consistent in this way because a person with the second-order intention necessarily works to be consistent in this way, not for consistency's sake but for the sake of right action.

Both these arguments rely on an analysis of the structure of right action that seems most faithful to ordinary language. Right action must be done *because* it is right, so it seems right action must be done out of the intention to do what is right. Yet to pursue this second-order intention seems to entail forming one's best judgment about what is right and trying to act on it. Intending the means one believes necessary to one's right ends makes up part of trying to act on one's best judgment. Working to the extent of one's abilities to eliminate inconsistencies in one's beliefs and act on the resulting picture of the facts make up two others. Someone with the second-order intention works toward states of consistency in her intentions and between her intentions and beliefs not for consistency's sake but because inconsistencies appear as mistakes from her perspective.

First, I set out the problem with the Instrumental Principle more fully. Second, I outline the above account of right action in further detail. Third, I begin the first account of the Instrumental Principle by arguing for the constitutive necessity of volitional means-end consistency to right action as such. Fourth, I explain how particular instances of means-end inconsistency give evidence of a general problem from the perspective of right action, namely a person's general inability or unwillingness to take necessary means. Fifth, I begin the second account by arguing for means-end inconsistency as a form of epistemic consistency and for a general requirement for agents to work to resolve inconsistencies in their beliefs. Finally, I employ the requirement to eliminate epistemic inconsistencies to produce the second account of the Instrumental Principle.

2. Why the Instrumental Principle Seems Right and Why It Seems Wrong

The Instrumental Principle does seem to have a good deal of ordinary language and practice in its favor. Someone declares her intention to become a doctor. However, when the time comes to apply to medical school, she does not do so. In such a situation, one might ask, “You know you have to apply to medical school before you can be a doctor, right?” If she were to respond in the affirmative, one could reply, “Shouldn’t you be filling out applications, then?” or “So don’t you think you should be filling out applications?” One could say these things even if one thought she had no talent for medicine and her gifts would better fit some other job. It seems one says both of these things with sense. But what sense is it?

As John Broome identified, assuming the existence of mind-independent facts about what is right, a requirement for persons always to make their intentions cohere in the way the Instrumental Principle prescribes would seem highly implausible, at least on its surface. On the one hand, making one’s intentions consistent does not seem like something one would have overwhelming reason in every case to do. On the other, it seems clear that following the principle would often produce wrong action. If one intends to do something that is not right, then intending the means to doing it would seem to cause one to intend yet another incorrect action. If it is not right for one to become a doctor then one probably should be using one’s time to do something other than applying to medical schools. If one intends even against one’s best judgment to satisfy one’s cravings for a cigarette, it seems unlikely smoking a cigarette would thereby become something one should do. So if it would lead one wrongly, in what sense “must” one’s intentions *always* follow the Instrumental Principle?¹⁹

Several attempts have been put forward to explain this “must.” For example, Broome himself argues one *does* have reason to be instrumentally consistent, but only in a “wide-scope”

¹⁹ John Broome. “Normative Requirements”. *Ratio* (1999) 12: 398–419.

sense of consistency. One has reason *either* to intend an end and its necessary means *or* to drop the end, not merely to do the former. Merely having ends does not necessarily entail one has reason to take the means.²⁰ For another, R. Jay Wallace argues that intending an end commits one to an incoherent view as of the facts about what is possible and impossible for one to accomplish. But one has reason to have an accurate view of the facts, which an incoherent view cannot be. So one should believe and intend consistently. I will return to this view at length shortly.²¹ For yet another, Joseph Raz argues that instrumental consistency is a kind of virtue. Persons who intend the means they correctly believe necessary to their ends in general succeed more frequently in doing the right thing than those who do not. Particular cases of instrumental inconsistency give evidence one does not possess that virtue. But one has reason to possess that virtue because it aids one to right action.²²

I wish to gainsay none of these views or any view hypothesizing or denying the existence of reasons to be instrumentally consistent. For all I will say here, particular reasons may or may not exist in favor of intending means one believes necessary to one's ends. There may or may not be a way in which consistent intention is also right action. I make no judgment as to that question. I want instead to suggest two other, different foundations exist for the Instrumental Principle. Both these foundations lie in the structure of right action as such. One foundation is the constitutive necessity of means-end consistent intention to any particular case of right action. The other is the constitutive necessity to right action as such of acting out of the second-order

²⁰ Most recently in John Broome. "Wide or Narrow Scope?" *Mind* (January 2007) 116: 359-70.

²¹ R. Jay Wallace. "Normativity, Commitment, and Instrumental Reason". *Philosopher's Imprint* (December 2001) 1 (3): 1-26.

²² Joseph Raz. *From Normativity to Responsibility*. Oxford University Press, 2012.

intention to right action. I will first outline this account of right action's structure and then go through these accounts in sequence.

3. An Account of the Structure of Right Action

First, though it is not universally agreed, it seems to have the weight of ordinary language on its side to hold one does not do what is right unless one does it *because* it is right. The character of this "because" is not universally clear, however. It seems most plausible that this is the "because" one uses in any other not-merely-causal explanation of actions. One saw or believed something was the case and determined oneself to act on that fact's basis. When someone asks why one went to medical school, one notes the features of attaining a medical degree in respect of which one determined to go there. Yet failing some deep materialist naturalism, those features did not exert a force that brought one passively into medical school. It seems doing φ because φ is right is the outcome of this same active, intentional determination rather than a passive, merely causal outcome. Doing the right thing for its own sake, then, would entail acting out of the intention to do what is right, a second-order intention guiding one's first-order intentions.

Second, it seems clear one cannot intend or act on the basis of facts one does not on any level recognize. One cannot stop at a stop sign "because" it is there if one is not in any way aware of its presence, though one may stop or be stopped accidentally where it stands. Nor does one do something "because" one was commanded to do if one never heard the command or if one did not hear it *as* a command. One cannot act or intend on the basis of facts one does not recognize any more than one can believe on the basis of evidence contained in books one has never read, in lectures one has never heard, or that one has not encountered oneself. To intend or

act in respect of p seems to constitutively require one to recognize that p . Thus the intention to do the right thing has as a constituent part the intention to recognize the facts about what is right.

Third, the intention to recognize the facts about what is right seems to have two constituent parts. As the first part, one intends to examine the facts themselves so as to determine what is right on the basis of the evidence rather than allow one's views to be determined by the causal force of one's psychology, the coercion of others, or the inertia of history. Second, one does this to the extent of one's ability. Everyone faces psychological, social, and historical obstacles. But persons face them with different abilities to overcome them and track the facts. In order to recognize the actual facts, then, each person must work as best she is able to overcome those obstacles. One has to work out the facts oneself to the best of one's ability. One has to form one's best judgment. This is the first part of the intention to do what is right. As the second part of that intention, one intends to act on that best judgment. Yet one does not intend to act on it *because* it is one's best judgment. One intends to act on it because if a judgment really is the best one can do, then from the first-person perspective, that best judgment as to the facts presents the facts themselves, so far as one is able to make them out. One forms one's best judgment in pursuit of an accurate view of the facts about what is right, as such. One acts on it in the attempt to do what is right.

This framework seems to show how one may be subject to requirements to form one's attitudes into certain patterns but not in virtue of any particular reason one has to do so. Right action as such requires the second-order intention to right action and yet the intention to right action constitutively necessitates forming the best judgment one is able and trying to act on it. The structure of right action itself constitutively requires one to intend a project that, if it succeeds even by one's own lights, necessarily results in the consistency of one's actions and

intentions with one's best judgment. One will not aim to form one's judgments into the best pattern one can thinking of it *as* a pattern or intend to act on one's judgments thinking of them *as* judgments. One will do these things in the attempt to do what is, in fact, right. Yet any particular right action must be performed out of that attempt to count as right action, at all.

The Instrumental Principle finds its basis in right action's constitutively requiring agents to form and act on their own best judgment about what is right because it is, they believe, right. I will argue that two aspects of this process require instrumental consistency. First, one must carry out an action in order to carry it out because it is right. This includes rightly achieving an end. But persons cannot achieve ends without taking the means they believe necessary to those ends, even if their actions causally produce states of affairs resembling their ends. Particular instances of the unwillingness or inability to intend means one believes necessary even to one's incorrect ends thus give evidence one lacks the general willingness or ability to bring about right ends, as well. Second, to do what is right, one must view it as right and perform it because it is right. If actual right ends count in favor of the actual means necessary to bring them about, I argue that sets of beliefs containing the rightness of an end but lacking the rightness of the means one believes necessary to that end are inconsistent. In the process of forming one's best judgment, one must try to eliminate that inconsistency and to take the means one views as necessary. But one must try to eliminate the inconsistency not for the sake of consistency but for the sake of doing the right thing. Inconsistencies appear from the first-person perspective not as internal relationships between beliefs but as one's own mistakes as to the facts themselves. Within the usual assumption of persons' generally pursuing courses of action they view as right, particular instances of instrumental inconsistency give evidence the inconsistent persons are not trying to do the right thing. But that intention is necessary to any right action, as such.

4. First Account of the Instrumental Principle

This argument has three main parts. First, I argue that persons who intend to do something necessarily *try* to take the means they believe necessary to their ends. Intention as such includes this constitutive part fundamentally because intention directs itself to the intention's success, that is, to *actual* action. Second, I argue succeeding in one's intentions partially constitutes achieving one's ends. One cannot achieve one's end by merely causally producing a state of affairs *resembling* the achievement of one's end. Third, I draw the conclusion from the previous two sections that intentionally achieving an end entails intending the means one believes necessary to that end. Fourth, I argue that this structure explains the universal applicability of the Instrumental Principle on account of what particular instances of instrumental inconsistency suggest about a person. One must intend with instrumental consistency to achieve *any* ends or carry out *any* actions, at all. But this includes right ends and right actions. Thus one who gives evidence of an unwillingness or inability to take the means she believes necessary to her ends, correct or incorrect, gives evidence of her unwillingness or inability to do what is necessary to accomplish right ends, as well.

4.1 The Attempt at Means-End Consistency as Partly Constitutive of Intention

My position as to the structure of intention and belief has precedent within recent literature. In "Normativity, Commitment, and Instrumental Reason," R. Jay Wallace suggests a person intending an end without intending the means she believes necessary to that end has an inconsistent set of beliefs. *Some* beliefs in an inconsistent set necessarily are false. Yet persons have reason to have true beliefs, he argues, and so they should believe and intend consistently.²³

²³ Wallace, pp. 14-26.

Wallace first recalls the nature of intention. Intending an end or action seems to entail one believes one may achieve that end or perform that action. A human may wish or want or imagine herself to fly to the moon on her own power, yet if she recognizes the laws of nature, she sees she has no chance of doing so. Thus she cannot intend to fly to the moon on her own power. This follows just from the nature of intention. In intention, one commits to *actually* doing something. But if one knows actually doing something is impossible, it seems one cannot intend it. What one will be doing constitutively cannot be intention.

But Wallace notes that consciously intending in instrumental inconsistency also involves one in an epistemic inconsistency. Say the following are all true. (1) One intends E. (2) One believes M-ing necessary to achieving E. (3) One does not intend to M. Premise (1) entails that (4) one believes it is possible for one to achieve E. But if one's belief (2) is true and (5) achieving E necessitates M-ing, and (3) one does not intend to M, then it is true that (6) it is impossible for one to achieve E. So at least if (7) one *knows* one does not intend to M, and whether or not one cares about it or becomes aware of it by following it out from what one believes, one's beliefs contain a contradiction. One cannot believe a means necessary to achieving an end, fail to intend the means one believes necessary to achieving that end, know one fails to intend that means, and *consistently* intend to achieve that end. Viewing oneself as failing to take a necessary means, one thereby (8) views one's end as impossible. Yet if one nonetheless intends the end, one thereby (9) views one's end as possible. Thus (10) one's views are contradictory.

Wallace suggests this inconsistency helps explain the problem with instrumental inconsistency. Inconsistent beliefs contain a falsehood somewhere. Yet it seems one has reason

to have true beliefs. So one should not, at least knowingly, intend inconsistently. It is right to have true beliefs, thus it is right to make one's beliefs and intentions consistent.

Joseph Raz, however, objects to epistemic-consistency-based accounts of the Instrumental Principle. Even if it is right for persons to have true beliefs, he argues, inconsistency only tells one that *some* belief in the inconsistent set is false. It does not tell one which it is. So even if one has reason to have true beliefs, why think one must always make one's beliefs consistent?²⁴

I want to argue in favor of Wallace's position that a requirement to pursue truth in one's beliefs about means, ends, and intentions grounds a form of the Instrumental Principle. But I do not argue this requirement applies in virtue of the reasons one has to have true beliefs or to make one's attitudes consistent. Rather, it seems a person with an intention necessarily *attempts* to bring consistency both (1) between her beliefs about what ends she may possibly achieve and her beliefs about the means necessary to her ends and (2) to her ends, her beliefs about the means necessary to her ends, and her volitional status with regard to that means. Yet she makes these attempts not because she has reason to do so but out of the constitutive structure of intention itself. I will argue a person who intends an end does (1) for the sake of determining the facts about the means sufficient for that end and does (2) for the sake of carrying out a sufficient means and that she does the former for the sake of the latter.

Persons intending to do something do not merely wish or want or imagine. They try to *do* something. Intention constitutively aims at action, at the intending person's actual efficacy within the world. A person who intends commits to *actually* doing something. Intention thus directs itself to the facts themselves, toward intentionally changing them. This is why, as Wallace points out, one cannot intend to do something one views as impossible.

²⁴ Raz, pp. 160-2.

Yet a person committing to actually doing something seems thereby also to commit to determining *how* she may do it. A person does not intend to actually do something who does not care to work out a method sufficient to actually carrying it out. This may not seem true at first. One does not have to deliberately work out how to carry out one's intentions most of the time. Most of the time, one does not doubt how to drink water, at least if one is socio-economically lucky. One goes to the faucet or the refrigerator without a second thought. One's world is familiar. Yet if one finds oneself in a desert, carrying out that intention becomes more complicated. One intends to drink water in a desert, in part, by seeking a method sufficient to find water and to drink it. If one does not try to form a plan as to how to actually find water and drink it, one seems not really to have the intention. One may have given up on finding water and may merely be dreaming about it. Yet in addition to working out a sufficient method, having the intention also clearly entails trying to actually carry out a method one views as sufficient. One may still fail to achieve the end. Nonetheless in taking a method one views as sufficient one thereby does what one can to achieve it. Intending, then, has two parts: (1) the attempt to determine a sufficient method and (2) the attempt to carry one out.

The structure of intention seems, then, to provide an answer to the question why persons must make consistent their beliefs about their possible ends and about the means necessary to those ends. It is not necessarily because of falsehood latent somewhere in inconsistent beliefs. Rather, it is because a person with an end necessarily seeks means sufficient to that end. She seeks a sufficient means until she finds one that, so far as she can tell, *does* suffice to the end, or else gives up the intention. She forms her best judgment as to the means sufficient to her end. "Best judgment" here designates not a particular judgment but the outcome of her attempt to find sufficient means. Yet the content of this judgment, *what* methods she judges sufficient, may

change over time. Her best judgment may remain fairly stable for awhile once she finds some water with no evident threats to her drinking from it when she sees fit. She must, however, *try* to change her views on the sufficient mean on learning new information. Say she believes she will drink water if she can just make it to the oasis-appearing object on the horizon. Yet say she discovers the appearance to be merely a mirage. One genuinely trying to drink water, thus genuinely trying to determine the means sufficient to drink water, necessarily will *try* to change her view of arriving at that object as a method sufficient to her acquiring and then drinking water. Sometimes this may be very hard or even impossible to do, as when one wants or needs for it not to be so. Yet a person attempting to drink water nonetheless must *try* to give that belief up. One “must” do so not because it is right to make one’s sufficient means-beliefs consistent with one’s other beliefs. One must do so as a constitutive part of trying to determine the facts about how to drink water, which itself partially constitutes the intention to drink water, as well. A person with an intention tries to make her views of the means sufficient to her end consistent not for consistency’s sake but for the sake of a clear view of the facts. But she seeks a clear view of the facts as part of her attempt at actual action, in which intention as such seems to consist.

This structure seems, in turn, to explain why one who views a means as *necessary* to her end must try to intend the means. The necessity of a means to an end conceptually entails the *insufficiency* of any method of achieving that end that does not include taking that means. Say it were the case one must dig into the earth to find any water in this desert, at all. Several different means of acquiring that water may still exist. One may dig with one’s hands or with a shovel or with a mechanical drill. Yet searching for an oasis or asking a person for directions to water will not, merely as such, suffice. Any way one goes, one must dig.

But why “must” a person who views a means as necessary to her end change her *views* about what means suffice for her ends? It seems the necessity of working out the facts to pursuing an intention produces an answer. A person with an intention must try to change her views about the methods sufficient to her ends to accord with her views about the means necessary to those ends because she constitutively must try to determine the *facts* about the means sufficient to her ends. She must try to accept that no means not including taking the means she believes necessary will suffice to her end, and thus that she cannot attain her end without taking that necessary means.

Take the following example. She sees an oasis on the horizon with no other water around. She believes she cannot make it past her current horizon and does not believe anyone will suddenly appear to give her water. So she believes making it to the oasis constitutes a necessary means to her drinking water. Now, if she intends to drink water, she constitutively tries to find a means sufficient to achieving that end. Yet now, by her own views, no method suffices to that end whereby she does not make it to that oasis. So she must *try* to accept that, say, heading back toward her town or following those wagon-ruts leading over the horizon will not allow her to drink water. She must do so as a constitutive part of trying to come to an accurate view of the means sufficient to her end. But she must do that, in turn, as a constitutive part of intending to drink water, of intending her end, at all. Indeed, it actually does not matter for the necessity of her working to bring this consistency to her beliefs whether her view about the necessary means is true. The facts, so far as she can make them out, are that no method suffices not including her going to that oasis. But inasmuch as she intends those ends she must work out the facts about the methods sufficient to her ends so far as she can make them out.

So, then, finally, one can say the following. A person who intends an end and believes by her best judgment in a means' necessity to that end constitutively must try to intend that necessary means. She must do this because of the following structures.

First, a person with an intention tries to determine the facts about the means sufficient to her end and tries to execute those sufficient means. She forms her best judgment about the means sufficient to her end and tries to act on it. She will, in other words, *try* to act consistently with her best judgment about the sufficient means. But she will do this not for the sake of consistency but for the sake of her intention's success, of actual action.

Yet, second, in the course of seeking a sufficient method, she may come to the conclusion that certain means are *necessary* to achieving an end. But if an end necessitates a means, that means also makes up part of any method sufficient to the end. It conditions the sufficiency of any means to the end.

Thus, third, a person with an intention must try to accept the insufficiency of any method not including the means she believes necessary to the end. She must try to make her beliefs about necessary and sufficient means consistent not, again, for the sake of consistency but as part of her attempt to attain an accurate view of the means sufficient to her intention's success.

Yet, finally, a person with an intention must also attempt to *carry out* a means she views as sufficient to her end. Yet if she has formed her views about the means sufficient consistent with her views about the means necessary, as she must try to do in forming her best judgment about the means sufficient, she necessarily *will* try to take the means she views as necessary. She will do so as a constitutive part of any method she views as sufficient to the task. She may go to the oasis by several different sufficient ways. But she believes that if she is going to drink water, she must make it there. So if she intends to drink water, she must set out for it.

Thus it is not necessarily *right* for a person to intend consistently with the Instrumental Principle. Nor is it that a person with an intention constitutively necessarily *does* take the means she views as necessary to her end. Rather, a person with an intention constitutively necessarily *tries* to take the means she views as necessary to her ends. She does this because she tries to employ a method sufficient to her end. Yet she must try to view any method not including the means she believes necessary as *insufficient* to that task.

So Wallace is right that persons must work for the sake of having true beliefs to bring consistency to their beliefs about their ends, the means to their ends, and their intentions. This may be, as he thinks, because persons just do have reason to have true beliefs. But whether they do have that reason or not, the following can be said for certain. Anyone with an intention must work to have true beliefs about those their ends and the means to their ends, and this will include working to have consistent beliefs about them. But they must work toward consistency because they must work toward an accurate view of how they may *succeed* in carrying out what they intend. They must work to have true beliefs about end and means because they must work to intend *effectively*.

A given instance of a person's failing to intend the means she views as necessary to her ends thus suggests various different conclusions. On the one hand, she may have tried and failed to accept the means' necessity to any sufficient method. Or perhaps she tried and failed to take those necessary means. Or maybe her best judgment was insufficient to accept there was no way to accomplish her end not including this means. Yet one explanation is available in any such case and indeed remains when any other fails. She may have failed to take the means she saw as necessary to her end because she was not *trying* to take the means she saw as necessary to her end. She saw what was necessary to accomplishing what she set out to do yet she either could

not or would not make the attempt to actually do it, thus to actually accomplish what she set out to do. On seeing what was required for succeeding in her aims, she either declines to or cannot undertake to do it, or else, since trying to take the means one sees as necessary to one's ends partially constitutes intention as such, she either gave up or never really had the intention.

These all seem to match the sort of diagnoses one would make in everyday life of evident cases of instrumental inconsistency. The Instrumental Principle would identify any such instances as problematic. But why should that be?

4.2 One Must Fulfill an End Intentionally to Fulfill the End at All

It seems the answer lies in the structure of action and thus of right action as such. To be unwilling or unable to do what one views as necessary to succeed in one's intentions is to be unable to succeed in one's intentions, including achieving one's ends, at all. But if one cannot succeed in intentions, one cannot succeed in *right* intentions, either. Particular instances of instrumental inconsistency give defeasible evidence one lacks the willingness or ability to act, thus also to act rightly. But that is a problematic lack in any case. This seems to follow, most fundamentally, from the nature of ends and thereby of right ends, as well.

An end seems uncontroversially distinct from what one sometimes calls an "end-state." End-states are merely the last in causal or temporal sequences of states of affairs. The end-state of the natural production of diamonds is the existence of diamonds. It is preceded by the co-presence of coal with extreme pressure. But when the coal and the pressure together produce diamonds they do not accomplish the end of producing diamonds. Coal and pressure cannot do so because, lacking intentions, they cannot have or accomplish ends, at all. Ends, unlike end-states, are the aims of persons. A person has an end when she commits to fulfilling some goal.

Her end is her project. She does not merely wish or suppose a state of affairs to occur. She commits to an actual accomplishment, to actually achieving the end. Ends, then, are not in the first instance ways for things to be but things for persons to do. Right ends, then, would be goals it is right for persons to achieve. An end could be right for a person to achieve even if she did not have the end. It could be right for many people to help reduce global warming even if they either disbelieve in global warming or do believe in it but do not see it as persons' responsibility to reduce it. One would act rightly in accomplishing right ends if one achieved the end because it was right to achieve it. By the above account of right action, that would mean achieving the end as the outcome of acting on one's best judgment that achieving it is right.

Achieving a right end seems a complex phenomenon, however. Say it is right to help reduce global warming. But say one knows nothing at all about global warming or the effects of fossil-fuel emissions on climatic temperatures. One replaces one's fossil-fuel car with an electrical hybrid. Yet one does this not to help reduce global warming but merely to raise one's social status within one's community. Replacing the car *does* lessen one's fossil-fuel emissions from their previous level, thus *does* help reduce global warming. Yet it seems clear one would not count as rightly achieving the end of helping reduce global warming. Why should that be?

The structure of right action, conjoined with the nature of ends, seems to give an illuminating answer. An end-state *resembling* one's achieving a right end occurred as a causal result of one's actions and intentions. Global warming was caused to slow some miniscule amount. Yet one did not undertake to help reduce global warming, nor did one actively take on a method adequate to accomplishing that end *as* a method for accomplishing that end. One undertook the action of buying and driving a new car as a method of accomplishing a different end, namely, raising one's social status. So though one's intention to gain status causally

produced a good end-state, that result was wholly accidental to the content of one's intention. But one cannot achieve ends, even right ends, one does not have. And where one does not achieve an end, at all, one cannot rightly achieve an end, either. What one does would not rightly be called "achieving" and the result would not rightly be called "an end." Global warming may have been reduced, but *one* did not reduce it.

The necessity of intention to the achievement of ends, including right ends, conjoined with the account above of intentional consistency's necessity to intentional success, thus seems to provide an explanation of why particular instances of intentional inconsistency are failures to do something one "should" do, regardless of the correctness or incorrectness of the end. The explanation is as follows.

4.3 Achieving an End Entails Achieving It with Instrumental Consistency

Section 4.1 suggested that one who intends forms her best judgment about a sufficient method for realizing her intention and tries to take it. This entails trying to take the means she views as necessary to her ends, thus to any method sufficient to her end. Section 4.2 suggested that achieving an end entails achieving it intentionally. One must accomplish the end *through* one's intentions. It cannot be merely accidental with respect to their content.

From these structures, three conclusions can be drawn about a person who achieves an end, correct or incorrect. (1) Any method she views as sufficient to achieving her end necessarily *will* include the means she believes necessary to the end. (2) She will have taken *one* of those methods she viewed as sufficient. But, then, finally, and most importantly for this investigation, (3) she *a fortiori* will have carried out the means she views as necessary to her end.

She will do so because that means will necessarily form a stage in *any* method she could view as sufficient to achieving her end.

Achieving an end of any kind thus entails intending in instrumental consistency. Ends must be achieved intentionally to be achieved, at all. A person intending an end constitutively attempts to intend the means she believes necessary to her end. She does so because she constitutively attempts to execute a method she views as sufficient to her end. She attempts to execute a method she views as sufficient to her end because she constitutively attempts to achieve her end in reality. Where her attempts to achieve the end succeed, she thus necessarily *will* intend consistently. Where she does not intend consistently, she does not succeed. The achievement is not hers.

This, I suggest, gives one account of why one finds instrumental consistency problematic in every case and thus one account of the Instrumental Principle. This then would complete one account of the Instrumental Principle.

4.4 The Instrumental Principle and the General Willingness or Ability to Do What is Necessary

Consider the painter Paul Gauguin. Gauguin abruptly left his wife and children to live in Tahiti for the sake of his art. Laying aside for the moment whether all right action is moral action, imagine Gauguin thought what he was doing was not right and assume further he was correct in that. Gauguin nonetheless committed to doing it and did it. He did so despite the risk of social opprobrium, ignominy, and failure.²⁵

²⁵ Bernard Williams considers Gauguin an exemplar of “moral luck,” whereby one commits to something that would be wrong if it does not work out but right if it does. In Gauguin’s case, in Williams’ estimation, the issue is whether Gauguin actually succeeds in producing great works of art. See Bernard Williams. “Moral Luck.” *Moral Luck*. Cambridge University Press, 1982. I use Gauguin as an example not of the relevance of actual outcomes to right action but of a person’s relation to her action in the moments of decision and execution.

Now consider Gauguin*. Gauguin* intends to leave his family and go to Tahiti to paint. However, standing in line to board the ship he believes he must board to make it to Tahiti, Gauguin* suddenly freezes up, turns around, and goes back to his family as the ship sails away. But he does not do this out of overwhelming guilt or the realization that life is more important than art. Instead, he is overcome merely causally by fear or anxiety so that he either cannot or will not get on the ship, that is, cannot or will not take the means he believes necessary to his end. He goes back to his family not because he realizes his mistake but because he cannot or will not do what he believes necessary to leave them.

Both Gauguins see what they set out to do as incorrect. Yet the Instrumental Principle identifies a fault in Gauguin* it does not find in the actual Gauguin. This fault applies or does not apply regardless of the right or wrong of Gauguin's ends. But why should that be? Why should it be, especially, since *ex hypothesi* it saved Gauguin* from making a mistake?

The answer seems to show up when one considers converse cases. Here is a set. Tom sits beside a busy sidewalk. He needs money and expresses that need to passers-by. Sara hears Tom and determines the right thing to do would be to give him some money. Say she is right in that. She sets out to give Tom some money. Now, there are various means sufficient for her to achieve her end. But say it seems to Sara all of them require her to approach Tom and extend the money. Sara sets out to achieve her end, say by the sufficient method of walking across the sidewalk and handing it to him. However, Tom exhibits behaviors indicative of mental illness, which causes Sara to fear him and makes her reluctant to approach. Yet say Sara overcomes her fear. She approaches Tom, hands him the money, and goes about her day.

Now suppose Sara* and Tom. Sara* also believes the right thing is to give Tom some money and that it is necessary to doing so that she approach him. Sara* sets out to give Tom

some money by the method she believes sufficient of walking across the sidewalk and handing it to him. Sara* also fears Tom. Yet Sara* does not overcome her fear, but gives in to it, abruptly stopping and turning to walk back the way she had come. However, unbeknownst to Sara*, she drops the money as she turns. Incognizant of having done so, she walks away hurriedly with the mere intention to exit the situation as quickly as possible. Tom, unaware of this entire sequence of events, finds the money on the sidewalk, keeps it, and uses it to buy the things he needs.

Both Sara and Sara* initially undertake a project sufficient to achieving her end. Both causally produce a state of affairs resembling the achievement of that end. Yet it seems clear Sara can be credited with achieving her end where Sara* cannot. Sara guides herself through a project sufficient to achieving her goal, which thereby includes taking the means she believes necessary to achieving that goal. She might still have failed. Tom could have refused to take the money or a passer-by could have grabbed it out of his hand. Yet these outcomes would result from way the world happened to go, not from any negligence on her own part. Sara*, on the other hand, declines to perform what she thinks is a necessary part of achieving her end. The course of events works in Tom's favor and Sara* unintentionally drops the money. Yet though Sara*'s actions and intentions play a causal role in bringing about that state of affairs, Sara* does not achieve the end. She did not execute an action she views as necessary to the end's achievement. Because she does not take the means she believes necessary to the end, she cannot intentionally carry out a project she views as sufficient to achieving the end. So she does not achieve it.

What is more, within the structure of right action outlined above, Sara acts rightly and Sara* does not. According to that structure, right action constitutively requires one to actively intend the action that is right and actually perform it because it is right. As has been argued

above, personally achieving an end requires personally taking and executing means sufficient to achieving that end, which in turn entails personally taking and executing the means one views as necessary to the end. Where one does not do any of these things, the achievement of the end does not occur under one's own control. But where one does not control an action, one cannot perform the action on the basis of its correctness, either. In other words, to perform an action on the basis of its correctness, the action must be one's own. But action being one's own constitutively requires one to take and execute means necessary and sufficient to carrying it out. Right action in part is autonomous action. But part of autonomous action is intending an action's necessary means.

Yet the question about the Instrumental Principle is the question why one may object to a person's failing to intend consistently in *any* particular case, even when her ends are incorrect. It has been assumed Gauguin* avoided doing something wrong in failing to intend the means necessary to going to Tahiti. The question is whether genuine ground exists for objecting to the inconsistency of his intentions, even so.

The answer seems to lie in what particular instances of instrumental inconsistency suggest about a person to an observer, even if the person and the observer are both oneself. Within the account of right action above, rightly bringing about an end entails having the end and intending the means necessary and sufficient to achieving that end because doing so is right. So doing the right thing in any particular case constitutively requires one to be both willing and able to undertake means necessary and sufficient to doing what is right. Where one cannot or will not do so, one does not do the right thing.

But, then, if acting rightly in a particular case constitutively requires one to take the means necessary to right action in that particular case, it would seem being apt to do the right

thing in general would constitutively require one to be in general willing and able to take the means one believes necessary to doing what is right. One cannot be relied on to do the right thing if one cannot be relied on to take the means one believes necessary to doing the right thing. Particular instances like Sara* declining to approach Tom give defeasible evidence of that general inability or unwillingness. Yet so does Gauguin*'s failure to board the ship he thought he must take to get to Tahiti. Right action makes up a mere sub-class of action more generally. One cannot achieve ends, at all, unless one takes means necessary and sufficient to achieving them. If one cannot or will not do what one's wrong ends require, one perhaps cannot or will not do what right ends require, either. Gauguin*'s declining to take the means he believed necessary to going to Tahiti may have saved him from doing something wrong in this particular case. But it nonetheless suggests one may not be able to count on Gauguin* to do the right thing if doing the right thing becomes equally as difficult as boarding that ship, as it sometimes does. In trying to determine on whom to rely in carrying out some project, one would not seek out the *merely* right-thinking. Right thinking does not always accompany the willingness and ability to do what is necessary, especially when it is hard. For right action, one can rely only on those who are apt both to determine what is right and actually do it. One needs companions who can think well about what is right but who also can and will do what is required to actually achieve it.

If one cannot or will not take the means one believes necessary to one's ends in any particular case, one gives evidence of one's inability or unwillingness to do the right thing, in general, as well. This evidentiary relationship applies regardless of the right or wrong of one's ends in the particular case. The failure to be instrumentally consistent in a particular case suggests an indisposition to actual right action, including achieving right ends. I suggest the

Instrumental Principle identifies this problem in Gauguin* and Sara*. Regardless of the right or wrong of their ends, it does not find it in Sara or in Gauguin as he actually seems to have been.

It is important that particular instances of instrumental inconsistency give only defeasible evidence of this problem. A person one thought unable or unwilling to do what is necessary on the basis of some set of performances may well exhibit great resolve, courage, and endurance in the next moment or when right action is actually at stake. One commonly makes these mistakes about people, even about oneself. Even so, regardless of the object, one's failure to form and maintain one's intentions gives evidence one will fail in the same way when genuine right action is at stake. If someone asserts the Instrumental Principle in the above sense, she expresses not necessarily a directly practical demand, but rather her doubts about one's abilities and dispositions. But those are abilities and dispositions one must have to be able and disposed to do the right thing. I argue that "must" explains the "should" of the Instrumental Principle.

This position shares much in common with Raz' views.²⁶ He agrees the requirement of instrumental consistency derives from the conditions enabling persons to carry out right action in general. Yet in his view consistency enables persons to do the right thing in only a probabilistic sense. Someone who takes or can take the means she believes necessary to her ends will be *more likely* to perform right actions than someone who does not or cannot. Instrumental consistency is useful for bring about right action and so one is to be respected for exhibiting it. Yet it neither suffices for acting rightly nor is it necessary to right action as such to do so. It is only one among several virtues rational persons may exhibit.

On the account I have tried to lay out, instrumental consistency is not just useful but constitutively necessary for right action. One cannot bring about right ends because they are right if one does not take the means one correctly believes necessary to those right ends. Thus

²⁶ Raz, pp. 157-60.

particular failures to take means believed necessary to one's ends give evidence of one's indisposition to right action. This view aligns with Raz in thinking of particular instances of instrumental inconsistency as evidential of a problematic character rather than incorrect action in themselves. Yet the character of which they give evidence seems more important than Raz' view would construe it. For Raz it is not *necessarily* a problem for someone to be unable or unwilling to take the means she believes necessary to her ends in general, even *most* of the time. Such a person can still perform right action. She just probably will not be very good at it and will perform some right action by accident. Within the account above, the general indisposition to intend means believed necessary entails the general indisposition to perform right action, as well. That indisposition is problematic in every case because one must do what is right in every case.

So this section's conclusions may be summarized as follows. On the one hand, Sara and Gauguin exhibit the same virtue. Sara is willing and able to do what is necessary for doing the right thing. She is able to do it even over the merely causal force of her subjective resistances to doing it. Gauguin is willing and able to do what is necessary for doing what will benefit his art. Even if one disagrees with what he does, one may admire, perhaps grudgingly, the courage with which he does it in risking public opprobrium and ignominy, besides failure. He does something wrong, yet one might say he does it in the right way. On the other hand, though good events occur from both Sara* and Gauguin*'s instrumental inconsistency, one nonetheless identifies the same fault in both of them. Sara*'s case seems to demonstrate *why* this is a fault. One can fail in right action with instrumental consistency. But one cannot succeed in right action without it. A person who acts rightly necessarily intends in accordance with the Instrumental Principle. Yet she does so not necessarily because she is trying to intend consistently or because she has reason to do so. She intends consistently because right action constitutively requires it.

Under this analysis, the complex sense of “You should intend M if you intend E and believe M is a necessary means to E” is: “One constitutively must take the means one believes necessary to right action in order to perform right action. To be relied upon to do the right thing, then, one must be relied upon to take the means one believes necessary to one’s ends. You intend E and believe M necessary to E. If you were in general able and willing to take the means you believe necessary to your ends, you could intend E now. If you fail to intend E, you give evidence either that either you cannot or you will not take the means necessary to your ends. Thus you give evidence you cannot or will not do what is right. A general disposition to right action constitutively requires one be in general willing and able to take means one believes necessary to one’s ends. In failing in this particular case, you give evidence you cannot be relied upon to do the right thing. But you must in every case do what is right.”

5. Second Account of the Instrumental Principle

5.1 The Instrumental Principle and the Transmission Principle

Under the account of right action outlined above, a necessary part of all right action is that one’s actions and thus intentions are consistent with one’s best judgment about what one should do. A necessary part of that requirement is, in turn, that one *forms* one’s best judgment. But one forms one’s best judgment when forms as accurate a view as of the facts as one is able. One does this by forming one’s beliefs as consistently with the evidence as one is able. One must do so not for consistency’s sake but in the course of one’s attempts to form a correct view of the facts. In turn, one must try to act on that belief not for consistency’s sake but because one is trying to do what is right. Right action requires both that one’s beliefs cohere in certain ways

and that one's intentions cohere with one's best beliefs. This structure, I will now argue, provides a second account of the "must" expressed in the Instrumental Principle.

In contrast to the Instrumental Principle, "Transmission Principles" do not prescribe certain patterns for persons' beliefs or intentions. Rather, they describe relationships existing between actually correct ends and the means that actually bring them about. These relationships are such that if an end E is defeasibly right, some means M that serves to bring about E about necessarily is defeasibly right. Raz, for instance, advocates for his "Facilitative Principle," which runs as follows:

when there is an undefeated reason to perform an action (the source action) there is also a reason to take any action which facilitates its performance, provided that it is part of a feasible and undefeated plan whose pursuit by the agent is likely to generate an opportunity to perform the source action, where a plan is defeated if the reason for any of its indispensable steps is defeated.²⁷

So where I should help the disadvantaged in distant countries, I genuinely have reason to give money to Oxfam or apply to work there or fly to those countries and hand money to people myself. These are matters of fact about how reasons and right action work.

For the purposes of this paper, it does not matter whether this or any Transmission Principle in particular is correct because it seems the following can be said with certainty. If *any* Transmission Principle is true, it must be the case that right ends count in favor of the means necessary to bring them about. It seems nonsensical to hold that right ends count in favor of means that are in some degree sufficient to bring about those ends but not to believe that right ends count in favor of the means without which there is no bringing-about of those ends, at all. If one asserts they do not, it seems one does not believe in Transmission Principles, at all. So if any Transmission Principle is true, it seems a right end counts in favor of its necessary means.

²⁷ Ibid. p. 148.

Transmission Principles have no direct consequences for the internal relationships between our intentions. They tell us about the internal relationships between the actions and states of affairs to which we must direct those intentions, not an order we must bring to our intentions independently of what is right. So at least on the surface of things, one very easily could do the right thing with respect to E and M but with means-end inconsistent intentions. Nonetheless, I want to say, if any Transmission Principle is true, then the Instrumental Principle has a second sense. Any beliefs whose contents violate a Transmission Principle are inconsistent. But agents must try to resolve inconsistencies in their beliefs to the extent they are aware of those inconsistencies and are able to do so. They “must” try, I will argue, as part of the structure of right action as such.

5.2 The Transmission Principle and Epistemic Inconsistencies

There are weaker and stronger forms of inconsistency between beliefs. A weak form is when the content of one belief is inconsistent with the evidence as represented in another belief. The fact that environmental science journals are reliable guides to the facts about global warming suggests that human-caused global warming is real. If I believe environmental science journals are reliable as to global warming but still believe that human-caused global warming is not real, then my beliefs are inconsistent. They are also inconsistent if I remain unopinionated as to the status of human-caused global warming. Yet evidential inconsistency is not the strongest form of inconsistency. It is still, for instance, conceptually and metaphysically possible that nearly every climate scientist has been congruently hallucinating for many years as to whether the global temperature data has shown a warming trend due to the rise of the use of fossil fuels. By contrast, if a Transmission Principle is correct, it is impossible for E to be correct, some M to be

a means necessary to E and M *not* to be defeasibly correct in virtue of those two facts. What is more, it seems clear that the kind of necessity encoded in Transmission Principles must be conceptual or constitutive necessity, such that E's counting in favor of some M follows just from E's being a right end. It does not seem to encode a necessity on the model of a physical law, such that a force other than E itself *causes* E to transmit correctness to M. Rather, it is contained merely in E's being a defeasibly right end and M's being necessary to E that M is defeasibly correct in virtue of that fact. The prior two facts in and of themselves alone thus would give deductive evidence of the third.

So it seems there are two possible scenarios in which belief sets would be inconsistent on account of a Transmission Principle. The first is that one believes E is correct, that some M is a means necessary to φ , and that M is *not* defeasibly correct. The second is that one believes E is correct and that M is a means necessary to E, but is unopinionated as to whether M is defeasibly correct. So if one believes one should be a doctor and that applying to medical school is necessary to becoming a doctor, but does not see applying to medical school as defeasibly correct in virtue of that relationship, one's beliefs are inconsistent.

This inconsistency, I now want to suggest, explains a requirement that persons make their beliefs about what is right means-end consistent. I will then argue that this requirement explains a second sense in which one "must" make one's intentions means-end consistent. I will do this by way of the requirement that agents form their best judgment of what is right as part of the second-order intention to do what is right.

5.3 Best Judgment and Consistent Beliefs

I have suggested above that right action must be done out of the second-order intention to do what is right for its own sake and that part of that intention is to form one's best judgment as to what is right. But forming one's best judgment is to form the most evidence-consistent belief one is able. For example, an investigator into global warming may be unable fully to understand the journals she reads or the lectures she hears or she may have incorrigible beliefs that force her to understand the evidence in a way compatible with those beliefs. So her attempts to follow the evidence as closely as possible may not produce in her exactly the same belief-states as would be produced in an expert working equally diligently with the same material. Nonetheless, the relation each person's best judgment would bear to her intentions, her abilities, and the evidence she recognizes would remain the same. No less with the expert than with the amateur, it would be the judgment that follows as closely as possible, given her abilities, the evidence she recognizes. What is more, once each arrives at her best judgment, she will not think of it *as* her best judgment or even think about it as a judgment, at all. From her perspective, it will be things as they really are as far as she can see them, viewed without the causal interference of her mental states or of any other condition. Yet from the third-person perspective, it will still be her best judgment, even if that best judgment is also true.

The requirement to work to eliminate inconsistencies in one's beliefs *as* falsehoods applies over all evidential inconsistencies. Where one is aware that one's views stray from the evidence, one must try and correct those views so as to bring them in line with the evidence one recognizes. One must in every case try to track the evidence so as to identify and do what is right. If one believes these environmental science journals are reliable but does not accept the

reality of global warming, one must *try* to accept it. If one does not make that attempt, one does not have the second-order intention.

Yet it seems if a Transmission Principle is true, any set of beliefs is inconsistent that includes (1) the correctness of E, (2) M's necessity to achieving E and (3) the incorrectness of M or (4) a failure to acknowledge M's rightness. Evidential facts *suggest* that something is the case. By Transmission Principles, an end's correctness *entails* the correctness of its necessary means. So from the perspective of an agent with one of these sets of beliefs, at least one of two things is the case. Either she is wrong in her view (3) that M is incorrect, or she has failed to draw the necessary consequence (5) that M is defeasibly correct, or both. From her own perspective, either one of her beliefs is false or she has failed to recognize the whole truth about what she should do, or both. So if she genuinely has the second-order intention to do what is right, she will attempt to give up (3) or accept (5), or both. This attempt partially constitutes the second-order intention, namely, forming her best judgment about what she should do. She must work to make her beliefs means-end consistent, but she must do this not for consistency's sake but as part of her attempt to make her beliefs true, to see things as they are.

It seems, moreover, the second-order intention requires one to try to bring a consistency to one's beliefs that is narrow in scope, not wide. One must not merely try to believe *either* that the means is correct *or* that the end is not. One must try to view as defeasibly correct the means to the ends one views as defeasibly correct. A person with the second-order intention to right action does not care about her internal consistency for its own sake. She cares about the facts and how best to respond to them. To work to see things as they are is to do one's best to accept whatever is the case even if it is discomfiting or counterintuitive. Mere wide-scope consistency does not fulfill but runs contrary to that project.

This is not at all to say one cannot view an end as correct, see the means that would be required to bring it about, and thus realize achieving the end would be wrong. It remains wholly possible within this account for one to view all the reasons weighing against taking on an end as outweighing whatever reasons count in the end's favor and to judge that, all things considered, the wrong of the means outweighs the right of the end. All that is argued here is that the outcome of one's best judgment cannot be to view an end as defeasibly right and not view that ostensible fact as counting in favor of its seemingly necessary means. The structure places no restrictions on the all-things-considered best judgments to which one can come. It only restricts the specific paths of reasoning one can take in arriving at those best judgments.

So now it seems one can make sense of the criticism to which one becomes susceptible when one fails to accept as right the means to the ends they view as right. A failure to be consistent in this way gives defeasible evidence one has not tried to determine the facts about what one should do and, in turn, that one does not intend to do what is right. There are many mitigating conditions that could dispel that explanation, such as one's lack of awareness of the inconsistency, one's inability fully to understand it, one's lack of time available for thinking through everything one believes, or one's lack of control over certain parts of one's psychology. Yet consciously to make no effort to accept what the evidence one acknowledges suggests when one is able to do so gives deductive evidence one is not trying to do the right thing.

5.4 Completing the Second Account of the Instrumental Principle

So far this argument only explains why persons must try to resolve means-end inconsistencies in their beliefs. The question as to the foundation of the Instrumental Principle asks why one must bring means-end consistency to one's intentions. Why must one intend the

means if one merely *intends* the end? Why must one do this even if one's ends are incorrect? It seems the necessity for agents to pursue epistemic means-end consistency gives a second account of the Instrumental Principle once an assumption within which we carry out the vast majority of our human social lives is accounted for.

It seems that in normal circumstances, when someone does something one does not understand or objects to, one views that person as appropriately subject to the question, "Why are you doing that?" When one's interest is not in causal antecedents, this is a request for justification. One wants the person to cite facts in virtue of which her action is correct, allowing one thereby either to approve or at least understand her actions. Viewing one another in normal circumstances as appropriate subjects of requests for justification would make no sense if we did not assume that people in normal circumstances do what they view as right.²⁸ A person who was not pursuing what she thought to be right could not answer one in earnest. Yet it seems it would be pointless for one to ask in earnest if one did not expect to be answered in earnest, if one expected everyone to lie or merely to express their non-rational psychological states in response. An earnest answer aims to show one why the person's action appears as correct from her own perspective and thus, as both one and the other person expects, will make it appear right or at least intelligible from one's own.²⁹ Asking people for the facts in virtue of which their actions

²⁸ It is worth emphasizing at this point that "right" and "correct" refer to the class including all those actions having decisive reason in their favor, or sufficient reason in cases in which no reasons are decisive. That we do not just ask for *some* reason but sufficient reason would seem to be suggested by our regarding it as open to us to continue the questioning until the other person gives us sufficient reasons or reasons that look sufficient to her, or else admits she did or intended it for no good reason, even from her own perspective.

²⁹ Of course the person could respond to us with the wrong kind of reason: (1) "Because I feel like it," (2) "Because it will help my career," (3) "Because it makes me feel morally superior to other people," (4) "Because I can't help it," and so on. It is imaginable that in certain circumstances a person could view these as genuinely justificatory facts. But usually they express that a person is *not* trying to do what is right and either (1) / (2) does not care or (3) / (4) knows this as a psychological fact about herself that she may or may not be comfortable with. These "answers" are not genuine answers but rejections of the question.

appear to them as correct would make no sense if one did not assume they were pursuing those actions in the view that they are right.

Within this assumption, two things can be expected from a person who has the second-order intention. First, she will do her best to accept the implications of her beliefs about what is right. One cannot respond rightly to circumstances if one has incorrect or limited conceptions of those circumstances, and so one is not trying to do the right thing if one refuses to accept the consequences of one's beliefs. Second, she can be expected to try to act on her completed view of what is right. She may not succeed in accepting the implications of her beliefs and various psychological obstacles or an emergency may keep her from acting on the particular view of right action she has at a particular time. Nonetheless, a person in normal circumstances will succeed in accepting the implications of her beliefs if she genuinely makes the attempt. So in normal circumstances one can expect a person to try to act on her beliefs' implications. One can expect her to form and try to act on a consistent view of what is right.

Now, as we have seen, a person believing E to be right and M to be a means necessary to E *must* try to believe that M is defeasibly correct. She must do so for two reasons that are ultimately the same. First, from her own perspective, her view of what is right would be mistaken or incomplete if she failed to believe in M's correctness. Second, trying to accept her beliefs' consequences is what a person with the second-order intention to do the right thing for its own sake necessarily *would* do and it is necessary to all right action that one act out of that second-order intention. To attempt to the best of one's ability to guide one's view of things by the evidence is what it is to try to determine what is right, which is necessary to trying to *do* the right thing for its own sake.

Yet the content of the second-order intention is not merely to determine what is right to the best of one's ability but to try to *do* it. So someone who believes E is right and M is necessary to E must not only accept M's correctness but also try to do M. If she intends E, she must try to intend M. She must try to intend M because that is what a person with the second-order intention does when she intends E. In normal circumstances, if one intends E and one believes M to be a means necessary to E, one must try to intend M.

If a person does not intend the means necessary to her ends, it may not be due to a lack of the second-order intention to do the right thing. She may not have been able to acknowledge M as right due to misunderstanding the concepts, her incorrigible attitudes, and so on. She likewise may not succeed in actually taking M. She may not fully understand M, her body may not respond or external circumstances may intervene. Her failure to *succeed* in believing the consequences of her beliefs or in taking the necessary means thus gives only defeasible evidence of her lack of the second-order intention. And yet a failure to *try* either to believe those consequences or to take those means gives *deductive* evidence of that lack. From her perspective, she refuses to accept the facts or actually to *do* what is right. But trying to do what is right so far as one is able to make it out is what the second-order intention *is*. A person who does not try to do either of those things is not trying to do the right thing, at all.

So within the surroundings of all these conditions, the "must" of "If you intend E, you must intend M" can have the following complex sense: "I presume you would not intend E if you did not think E was right. But you cannot E without M-ing and ends count in favor of their necessary means. Thus if you are genuinely trying to do the right thing, you must try to accept that M is right and you must try to do it. Otherwise you lack the second-order intention. But

inasmuch as there is a right thing to do, you must have the second-order intention to do the right thing for its own sake and so must try to view M as right and intend M.”

In other words, when one finds the person who has declared her intention to go to medical school not filling out applications in the run-up to the deadline, one of the things one can say to her is, “Don’t you think you should be filling out applications right now?” When one says this, one expresses both parts of this account of the Instrumental Principle simultaneously. One calls her to action, yet one issues that call by demanding that she face up to the implications of the beliefs one assumes she has. If, as one assumes, she thinks this end is right, then if she is serious about bringing it about, she must try to accept that the necessary means are required of her, too, and must try to take those means. She “must” try to accept the implications and intend the means because that is what someone with the second-order intention necessarily does.

This is a second sense for the Instrumental Principle. One clearly would use it in contexts very different from those in which one would employ the first sense. It would communicate a different criticism. In the first, one impugns a person’s ability or willingness to carry out right actions. In the second, one questions her intentions, regardless of her ability to carry them out. One can apply either criticism to oneself as to another without loss of sense.

6. Conclusion

The question about the Instrumental Principle is the question as to the sense of “should” one uses when one tells a person that if she intends an end, she “should” intend the necessary means to that end, even if one does not agree she should pursue the end and does not view the means as worth doing in itself. How could a person be “required” to do something that is not right so as to do something *else* that is not right?

I have tried to identify two such senses. In the first, someone says one “should” take means one believes necessary to the end one intends because intending the means believed necessary partially constitutes achieving an end because it is right. One cannot achieve an end unless one achieves it intentionally. But intending an end entails trying to determine a means sufficient to the end and carry out that means. Yet trying to determine a means sufficient to an end entails trying to accept that no means is sufficient that does not include the means one views as necessary to that end. Where one succeeds in achieving an end, then, one does so by successfully taking a means one views as sufficient, which must include whatever means one views as necessary. Where one succeeds, one’s intentions thus *will* be means-end consistent. This includes succeeding in achieving an end that is right. The application of the “should” of the Instrumental Principle to cases of incorrect action derives from what particular instances of instrumental inconsistency suggest about one, regardless of whether one’s end is correct or incorrect. Being in general able and willing to do the right thing constitutively requires one to be in general able and willing to take the means one believes necessary to one’s ends. Particular cases of instrumental inconsistency thus give defeasible evidence of one’s general indisposition to right action. The former are defeasibly problematic in every case by giving evidence of the latter. The latter is indefeasibly problematic in every case because one must do what is right.

In the second sense, one says a person “should” intend the means necessary to her ends because that is what a person necessarily *would* do who tried to do the right thing and genuinely believed her end to be right. One’s general presumption that persons are in general subject to questioning as to the justification of their actions would make no sense if one did not assume that persons in general take on actions because they view them as right. Within this presumption, one can expect an agent with an end to do two things. First, one can expect her to try to view the

means necessary to that end as right. One can expect this if any Transmission Principle is true and right ends necessarily count in favor of their necessary means. To fail to attempt this would be for the agent, from her own perspective, to acquiesce in an inaccurate or incomplete view of the facts. Second, because of the first expectation, one can expect the agent to intend the means. To fail to do so would be for the agent to acquiesce, from her own perspective, in *not* doing what is right. One can expect both these attempts because a person with the second-order intention to do the right thing for its own sake necessarily *would* do these things. That second-order intention has as its constituent parts the formation of one's best judgment as to what is right and the attempt to act on that best judgment. A person trying to do the right thing cannot consciously have an incorrect or incomplete view of the facts. Neither can she decline to do what she views as right. So inasmuch as one can expect people to have the second-order intention and inasmuch as one assumes them to intend the ends that they believe to be right, one can expect them to intend the means they believe necessary to their ends. But one can expect persons to have the second-order intention because it seems a necessary constituent of right action as such.

So under both of these accounts, the “must” of the Instrumental Principle is ultimately the “must” of the constitutive structure of right action. A person “must” be consistent in these ways or else give evidence she lacks the states or abilities right action as such requires. On the one hand, “one must intend M if one intends E and believes M necessary to E” can mean that right action requires one to be able and willing to intend M when E is right. The speaker suspects one lacks that ability and willingness. On the other, it can mean that, if one had the second-order intention, which one must have to do the right thing, one necessarily *would* intend M if one believed E to be right, as the speaker presumes one does. The speaker suspects one may lack that intention. In both cases, she looks for evidence to the contrary.

CHAPTER 4

Right Action and Integrity

1. Introduction

Say there is a right thing to do. Say the right thing to do is defined in every case as the action that brings about the optimal state of affairs. In philosophy this view about the content of right action is called “consequentialism.” When the question arises as to the right thing to do, consequentialism gives a single answer in every case. Whatever action produces the state of affairs of the greatest value is the right action.

Consequentialist conceptions of right action as such entail the following. Circumstances can arise in which doing the right thing requires infidelity to some of the commitments central to one’s life. It is often taken that one’s primary responsibility is to one’s family. One’s resources should go first of all to ensuring their welfare and only secondarily to friends or strangers. Some view this order of priority as holding even if one is already well-off and puts those resources, say, to sending one’s children to private rather than public schools. But if, say, the amount of human happiness achieved determines the value of states of affairs, saving a stranger’s children from starvation would be a better action than sending one’s well-off child to a private school. So under this consequentialist theory of value, other things equal, a well-off parent should give those resources to the stranger’s children instead of her own. The consequentialist structure of the theory, not merely the particular conception of value, entails this result.

A pure consequentialism evaluates different courses of action merely quantitatively: how good a state of affairs it would produce. Merely as such, a consequentialism has no principled way to assign a conditioning role to one’s relationships, goals, religious views, one’s personal commitments as such, in determining what one should do or when one acts rightly. Since all

evaluation is in terms of the quantitative value of states of affairs, the qualitative differences between actions can receive no special significance merely as such.

Bernard Williams takes these facts to entail that pure consequentialism fails as a complete theory of right action's structure. Intuitively, he argues, it *cannot* be the case that a person must act against her own deepest commitments, the projects or attitudes according to which she structures her life, in order to do the right thing. "It is absurd," he says, "to demand of a man, when the sums come in from the [value-creation] network ... that he should just step aside from his own project and decision and acknowledge the decision which [consequentialist] calculation requires."³⁰ In his view, any correct theory of right action must preserve the role of what he calls "integrity," a person's fidelity to her commitments, in conditioning the content of right action. In certain circumstances pure consequentialism would require one to contravene one's integrity in order to do the right thing. And it would do so without there being a further meaningful question to ask beyond the value of the consequences as to whether one should do it or would act rightly in doing it. Thus pure consequentialism must be false.

In this paper I want to suggest that under a conception of "integrity" closely tracking the ordinary meaning and evaluative relevance of that word, this argument is not intuitive but deductive. Integrity not merely seems to condition any possibly true account of right action. I will argue the structure of right action as such entails it constitutively must do so. Right action must be done because it is right, which seems to entail that it must be done out of the intention to do what is right because it is right. But to intend to do the right thing because it is right is to form one's best judgment about what is right and try to act on that best judgment. Right action must be performed as the outcome of forming and trying to act on one's best judgment about

³⁰ Bernard Williams. *Utilitarianism: For and Against*, with J.J.C. Smart, Cambridge University Press, 1973, p. 116. Williams has the "utility" network and "utilitarian calculation," but his critique is directed against the consequentialist structure of utilitarianism, not the utility criterion for right action's content, in particular.

what is right. But it seems forming and trying to act on one's best judgment about what is right in every case is or may be the core sense of what we mean by "integrity."

Within this structure, a person does not count as acting rightly in a particular case if she does not act on her best judgment about what is right in doing so. Yet this is not to say the action in question could not be correct in other circumstances. However, I will argue, integrity also conditions the actual content of right action as such. Certain actions cannot be right in principle, are incorrect in virtue of their structure alone, because they require the contravention of one's integrity as such. Actions constitutively requiring one not to form or try to act on one's best judgment constitutively cannot be right action, at all. Pure consequentialisms would require one to undertake such actions in certain circumstances. Pure consequentialisms thus cannot give the content of right action. In this way integrity must condition persons' first-personal deliberation about what they should do as well as the third-personal evaluation of their performances.

First, I set out Williams' argument against consequentialism in more detail. Second, I set out two desiderata for any adequate account of integrity. It must mark important distinctions in the evaluation of action and must suffice for ordinary uses of the word "integrity." Third, I give the analysis of right action as the outcome of forming and trying to act on one's best judgment about what is right. Fourth, I show how attempting to act on one's best judgment about what is right satisfies the desiderata for an account of integrity. Fifth, I argue that the structure of right action as entailing action with integrity makes sense of the conditioning role Williams suggests integrity must play in acting rightly in any particular instance. Sixth, I suggest this structure shows how integrity conditions right action's content as such. Finally, I show that the structure of right action entails that action constitutively requiring one not to form or act on one's highest-

order best judgment about what is right cannot be right action. There is nothing it is like to do it for its own sake, and so forfeiting one's integrity as such can never be right.

2. Williams against Consequentialism

In arguing for the importance of "integrity" Williams gives the following examples:

- (1) George needs a job very badly so as to take care of his family. The only one he can get is at a laboratory researching chemical and biological warfare. If George does not take the job, it will be taken by someone very zealous for it who would create more effective implements more quickly than would George. However, George is opposed to chemical and biological warfare.
- (2) While visiting another country Jim runs upon a detachment of government soldiers holding twenty villagers at gunpoint. Their captain tells Jim that if Jim shoots one of the villagers he will let the other nineteen go free. If Jim does not, the captain will kill them all. The villagers against the wall themselves ask Jim to do it. But Jim believes it would be wrong.³¹

A pure consequentialist theory wherein the aggregate human happiness achieved in states of affairs determines their value would seem to entail that George should take the job and Jim should shoot one of the villagers. A consequentialism of this kind views these as *obviously* the right answers. No further question beyond the quantity of value generated makes sense to ask.

Williams does not direct his argument against these answers as such. In fact, he thinks that in Jim's case a consequentialism of this kind probably gets the right answer.³² He argues instead against the *obviousness* of these answers. Some further question must remain to be asked beyond the value of the state of affairs the action would produce. This further question must condition the correct answers about right action in a way different from the calculus of the value of states of affairs.

³¹ Ibid. p. 98. The condition that Jim believes it is wrong to kill human beings is implicit in Williams' initial presentation, but necessary for the integrity objection to make sense in this case. At least on the surface of it, if Jim were a utilitarian, there would be no problem as to Jim's integrity here, at all.

³² Ibid. p. 117

Williams presents an action's relation to a person's integrity as his diagnosis of the further question to be asked. It may be right on the whole for George to take a job developing biological and chemical weapons so as to take care of his family and hold back the development of those weapons and for Jim to shoot one villager to save nineteen rather than allow all twenty to be killed, even if either person thinks it is wrong. But George and Jim's convictions against doing those things must play *some* role in determining when they act rightly and what it is right for them to do. An action's relation to a person's fidelity to her own commitments, "the distinction between my projects and someone else's projects," must partially determine whether she would act rightly in doing it.³³

Consequentialists do take *some* account of a person's commitments, convictions, relationships, values, projects, concerns, preferences, and so on. Other things equal, requiring infidelity to one's commitments could reduce a state of affairs' value below the value of an alternative state of affairs not requiring it. But other factors, for instance the lives of twenty people, may always outweigh that decrease in value. Williams' intuition suggests integrity's role in determining the content of right action must be much stronger than this. Even if a person acting against her commitments would bring about an ideal state of affairs, Williams thinks one must be able to meaningfully ask whether *she* should be the one to do it. The relation of fit or lack of fit between her commitments and an action must somehow *condition* whether she would act rightly in doing it and what it is right for her to do. Pure consequentialism could make no sense of that determining role. So pure consequentialist theories must not give correct accounts of the content of right action.

³³ Ibid. p. 117

3. Desiderata for an Adequate Conception of Integrity

In trying to elucidate the role of integrity with reference to acting rightly and the content of right action, it seems the conception of integrity should meet two conditions. First, it should make evaluatively relevant distinctions among cases. Second, it should track ordinary uses.

3.1 Making Relevant Distinctions

First, it seems not *all* deep commitments should matter to the content of right action. Persons enter into relationships, convictions, interests, preferences, and so on in different ways and for different reasons. A person's prejudiced convictions would not seem to determine whether she should do something in the same way as convictions at which she arrived by conscientious consideration would. Nor would the relationships one enters into through a psychological compulsion seem to play that role in the same way as would those one views as objectively one's highest responsibility. Nor would one's obsessive interests seem to play that role just as would those interests one has because they seem to one to be genuinely important. If one cannot make these distinctions, it seems one cannot distinguish between the sorts of "integrity" that determine right action's content and which do not.

Consider the following modulations of Jim's case. Let me state emphatically that all these cases operate according to the assumption that the consequentialist answer is the right one. My own views about what is right do not play a role in this discussion.

- (1) Jim is reluctant to shoot anyone because he believes it is wrong and refuses to consider the alternative out of a desire to remain comfortably assured of his moral purity within a moral worldview he is comfortable with.
- (2) Jim is reluctant to shoot anyone because he believes it is wrong and will not consider the alternative out of fear of what the right action might require him to do.
- (3) Jim is reluctant to shoot anyone because it is wrong as far as he is able to tell by the most rigorous judgment of which he is capable.

It seems one would be disinclined to think Jim's commitments should determine what he should do in all of these cases. Jim's reluctance arises from a belief with the same content in every case. Yet Jim comes to or maintains that belief in very different ways. In the first two, whether overtly or covertly, Jim holds to the belief because it serves one of his interests or as the causal result of a psychological state. But in the last case Jim holds to the belief only because, so far as he can tell, it really is true. The facts, as far as Jim can make them out, are that he should not shoot anyone. This distinction would seem to make an important difference to whether Jim would act rightly in contravening his belief.

3.2 Relation to Ordinary Language

Second, when would one say a person has or acts with integrity? It seems one would say this of someone who, for instance,

- (1) does not exploit situations to her own interest when she could, or
- (2) does the right thing even when outward circumstances make it difficult, or
- (3) does the right thing even when she is not inclined to it, or
- (4) faces the facts even when they are contrary to her interests, or
- (5) is conscientious in her evaluation of evidence, even if not always accurate, or
- (6) changes her beliefs or actions when she realizes they are mistaken, or
- (7) is careful to do the right thing completely, or
- (8) keeps her promises, or
- (9) does not lie, or
- (10) genuinely tries to do any of these things, where trying to do them but failing makes sense, as it would not seem to under (1) or (9),³⁴

By contrast, it seems one would *not* say someone has or acts with integrity if, say,

- (11) she holds to her commitments despite recognizing them to be founded on false beliefs, or
- (12) she holds to her commitments merely out of a psychological compulsion, or
- (13) she holds to her commitments merely because they serve some interest of hers, or
- (14) she acts or intends on the basis of commitments formed and held to in any of these ways, or

³⁴ However, if someone *never* succeeded in any of these intentions, one would not be inclined to call her a person of integrity. We assume the world does not frustrate our plans to such an extent.

- (15) she holds to some commitment but carries it out carelessly, and
- (16) she makes no genuine effort to change any instance of (11)-(15), or else to make up somehow for the inability to do so.³⁵

In cases (11)-(16), one might say of a person that she had or acted with “a *certain* integrity.” But one would indicate one was speaking in an extended sense in using that locution.

Say one conceived integrity as simple congruity between one’s commitments and one’s actions or intentions without thought for how those commitments were formed. On the one hand, this conception would recognize cases (14)-(16) as instances of integrity and could not make sense of why (11)-(13) cases would rule out that attribution. On the other hand, it could not readily account for why one *would* say a person had or acted with integrity in cases (1)-(10). Such a conception thus would seem too broad to capture ordinary uses of the term. An adequate conception would need to be sufficient to them all.

I will try to formulate an adequate conception of integrity in the following way. It seems a conception of integrity more adequate both to ordinary language and intuitive judgment partially constitutes right action itself as ordinarily conceived. The structure of right action itself seems to constitutively necessitate that persons form and try to act on their best judgments about what is right in every case. But forming and trying to act on one’s best judgment about what is right seems to be what is meant by “integrity” in most ordinary contexts and seems to mark evaluatively important distinctions in the different ways persons can succeed or fail in living up to their own commitments. Yet beyond providing a more or less sufficient analysis of the concept of integrity, this constitutive structure seems in turn to provide a systematic account of the role Williams thinks persons’ integrity must play in determining when they act rightly as well as the content of right action, as such. I will argue that, in virtue of right action’s constitutive

³⁵ Though, again, if one never succeeded in doing otherwise, one would not refer to that person as acting with or being a person of integrity. Again, we assume persons’ psychology is not that intractable.

structure, action accidentally requiring one to contravene one's integrity cannot count as acting rightly and that actions constitutively requiring one not to act with integrity cannot count among the possible contents of right action, at all.

4. Right Action

What is it to do the right thing? It seems to have the weight of ordinary language on its side to hold that one does not do the right thing unless one does it *because* it is right. Yet the character of this "because" is not universally clear. It seems most plausible that this is the "because" ones use in any other non-causal explanation of action. One saw or believed that something was the case and actively determined oneself to act on that fact's basis. When people ask why one went into one's career, one notes the features of that career in respect of which one actively determined to pursue it. Failing some deep materialist naturalism, those features did not exert a force that brought one passively into that course of action. It seems doing ϕ because ϕ is right is the outcome of this same active, intentional determination, rather than a passive, merely causal outcome. To do the right thing, then, one must act out of the second-order intention to do what is right because it is right.

It seems clear one cannot intend or act on the basis of facts one does not in any way recognize. Though one may stop or be stopped accidentally where it stands, one cannot stop at a stop sign "because" there is a stop sign there if one is not in any way aware of the presence of the stop sign. Nor does one do something "because" one was commanded to do it if one never heard the command or if one did not hear it *as* a command. It seems one cannot act or intend on the basis of facts one does not recognize any more than one can believe p on the basis of evidence contained in books one has never read, expressed in lectures one has never heard, or that one has

not encountered oneself. To intend or act on the basis of the fact that p seems to require one to recognize the fact that p . Thus the intention to recognize the facts about what is right partially constitutes the intention to do the right thing for its own sake.

To intend to recognize the facts about what is right seems in turn to have two constituent parts. In the first part, one intends to examine the facts themselves in order to determine what is right on the basis of the evidence rather than allowing one's views to be determined by causal forces. And further, one does this to the extent of one's ability. Everyone faces causal obstacles to obtaining a correct view of the facts by examining the evidence, whether psychological, physical, social or historical. But persons face them with different degrees of the ability to overcome them and track the facts as they are. In order to recognize the actual facts, one must work to the extent of one's abilities to overcome those obstacles. One has to form one's best judgment. As the second part of the intention to do the right thing for its own sake, one intends to act on that best judgment. But one does not intend to act on one's best judgment viewing it *as* one's best judgment. One intends to act on it because, from the first-person perspective, it presents the facts about what is right as far as one is able to make them out. One forms one's best judgment in the pursuit of an accurate view of the facts about what is right and one acts on it in the attempt to do what is right for its own sake.

So the intention to act on one's best judgment about what is right partially constitutes all right action, as such. One must do the right thing out of the attempt to act on one's best judgment in order actually to do the right thing. But this means all right action must be performed out of a narrow form of integrity. It is not that one must act in fidelity to one's commitments, however formed. Rather, one must act in fidelity to one's best judgment about

what is right. But to act on one's best judgment about what is right because it is, one believes, right, one must act out of the intention to do the right thing for its own sake.

I now want to suggest three things. First, trying to act on one's best judgment makes what seem to be evaluatively important distinctions between cases. Second, it seems almost entirely adequate to ordinary uses of the word. Third, this account of right action's structure gives a systematic account of the conditioning role Williams argued integrity must have in determining both when persons act rightly and the content of right action.

5. Satisfying the Two Desiderata

5.1 Making Relevant Distinctions

This narrow conception of integrity seems to make the evaluatively relevant distinction between Jim (1) and Jim (2) on the one hand and Jim (3) on the other. It also tells us *why* it is important. In cases (1) and (2), Jim declines to shoot someone not because, so far as he can tell, it would be wrong to do so. Deep down he believes what he believes because it serves some interest of his or out of a psychological compulsion against not believing it. He does have a view as to the facts about what he should do. Yet he does not view those facts *as* the facts, but as what he needs to believe in order to accomplish a purpose of his or to avoid something he fears. He could ask himself a further question but declines to do so. So if Jim (1) or Jim (2) were to act on his commitments, he would not count as acting with integrity. Yet if he were to act against them, he would not count as acting against his integrity. Thus the relation his actions would bear to his commitments would not be relevant to the evaluation of his actions, at least within the above account of right action's structure.

By contrast, Jim (3) forms his belief as the outcome of his best judgment. He views it as a fact that he should not shoot a villager because it *is* a fact as far as he can make out. He cannot by further effort see *around* that view. Unlike the others, he takes on his commitment because it is, so far as he can tell, the commitment he objectively should have. So if he acted on that commitment he would count as acting with integrity, and if he acted against it he would count as contravening his integrity. Both facts would bear on the evaluation of his actions within this account of the structure of right action. This conception of integrity thus seems to make an evaluatively relevant distinction between cases of acting on or contravening one's commitments and to make that distinction in the right place.

5.2 Relation to Ordinary Language

Likewise, as an account of the ordinary concept of integrity, it seems the conception of integrity as forming and trying to act on one's best judgment about what is right for its own sake succeeds almost completely. Integrity as trying to act on one's best judgment about what is right for its own sake unifies the quality one recognizes in a person when she exhibits all of the various ways of thinking, intending and acting described in cases (1) through (7) and (10). These cases differ radically as outward behaviors, yet one identifies the same quality of character, integrity, through them all. This account clarifies why one would call all these outwardly different behaviors by the same name. A person who tries to do the right thing because it is right necessarily will try to do *all* of these things. This account does not yet make sense of cases (8) and (9). It seems likely, however, that those cases may also be derived from this conception of

integrity once conjoined with a truth-norm and a requirement to form and maintain cooperative projects with others.³⁶

Likewise, this conception of integrity would rule out cases (11)-(15), meeting condition (16), of persons *not* acting with or being persons of integrity. Forming one's best judgment means precisely to work to the extent of one's ability *not* to believe things because one wants or is compelled to or because one refuses to consider the alternative. It also means trying to act on that best judgment, which entails fulfilling its contents' requirements completely.

6. Integrity and Acting Rightly

Within the structure of right action outlined above and with the conception of integrity refined to make relevant distinctions and better reflect ordinary language, it seems one can now give a systematic account of Williams' original intuition that a person's integrity *must* determine when she acts rightly in a way qualitatively different from the quantities of value her actions would produce.

Say Williams' view about what Jim should do is false. Jim in fact should *not* shoot anyone. Jim (1)-(3) has the *correct* view about the right thing to do. But now say all of them act on that belief. In this scenario, Jim (3) would count as acting rightly. He has formed his best judgment about what is right and acts on it because it is, he believes, right. As it turns out, he is correct. Thus he does the right thing because it is right. Jim (1) and Jim (2), however, have not formed the best judgment of which they were capable about is right. As a result, their own view as of the facts about what is right does not appear to them *as* presenting the facts about what is right. They view it as the view about what is right they are comfortable having or that they are afraid not to have. There is, as it were, a second-order barrier between themselves and the world

³⁶ I hope to produce this account in the near future.

as they see it. They cannot do what is right because it is right because they do not view the facts about what is right *as* the facts about what is right. So they do not act rightly, even if they perform the correct behaviors.

On the other hand, say Williams' view about the right answer is accurate. Jim (1)-(3) has it wrong. He should shoot someone. But say he declines to do so on the basis of his belief that it is wrong. He would fail to do the right thing in every case. But the *way* he would fail would be importantly different. Jim (3) would fail because he tried and failed to do the right thing. Jims (1) and (2) would fail to do the right thing because they did not try. Acting with integrity is a necessary though not sufficient condition of acting rightly. Persons who form their best judgment and try to act on it still may be mistaken about what is right and so may fail to do or intend the right thing. But nonetheless they fulfill a condition without which no right action is possible. Conversely, persons who either do not form their best judgment about what is right or do not try to act on it contravene that condition. Thus they do not act rightly, even in performing an otherwise correct behavior, regardless of what they go on to do.

This difference in the way a person can fail to do the right thing seems to coincide with Williams' insights about the way consequentialism reduces agents to mere loci for the production of good states of affairs.³⁷ On this account, the structure of right action itself entails the significance of persons' commitments and intentions in every case. In every instance of right action, there must also be an agent performing the right action out of her own view about her commitments and her own intention to fulfill them. She cannot merely be an item in a larger causal trajectory that happens to result in better states of affairs.

On the other hand, the account also explains how persons can believe and intend, one might say, in the right *way*, even if the mind-independent facts about what is right differ from

³⁷ Ibid. pp. 116, 118.

those beliefs. Jim (3) did not decline to ask himself difficult questions out of a contrary interest or fear. He guided himself as best he could toward the facts *as* the facts. In a way it was merely an accident he did *not* do the right thing. He did everything in his power to get it right. By contrast, for Jim (1) and Jim (2) it was not an accident. Jim (1) and Jim (2) decline to investigate the question of what was right as far as they were able. There was something more they could have done to get the right answer. Jim could have faced a hard question for his moral worldview or overcome his fear, or at least tried to do either of those things. But he did neither of them. By contrast, from his own perspective, when Jim declines to ask himself a question out of discomfort or fear he declines to determine the *objectively* right thing to do. He thus leaves it to chance whether what he wants to believe or fears not to believe will turn out to be true. If it turns out they should not shoot anyone, Jim (1) and Jim (2) are lucky. But the credit for their performance of the correct behaviors goes not to them but to the way the world turned out. If it turns out they should, they are unlucky, but the failure to do the right thing is all their own. Yet either way the facts turn out, Jim (3) does all he can. Thus if Williams is wrong, Jim (3) shares the credit for his right action with way the world turns out. If Williams is right, Jim's failure is all on the world. Jim himself acted with integrity, either way.

Even if Jim (3) is mistaken, he would seem to gain one's respect. One would seem to accord him something like the respect with which one regards, say, Ptolemaic astronomers who worked with absolute diligence to construct cosmological systems that turned out to be radically false or like those who live conscientiously according to values one does not share. Like the latter, even if Jim (3) hits on the wrong answer, he organizes himself in the way necessary to right action to such an extent that it will be an accident if he goes wrong. By contrast, one would not view Jim (1) or Jim (2) the same way even if they accidentally behave correctly. So it would

seem this conception of integrity both follows from the ordinary concept of right action and suffices for one's ordinary employment of both those concepts.

Williams' concern was in part to establish that a person cannot *simply* count as acting rightly in a particular case if her actions produce the best state of affairs. Whether she acts out of or against her integrity must partially condition the answer to that question. The account outlined above makes systematic sense of that intuition. The other part of Williams' concern, that integrity must partially condition the contents of right action, still remains to be explained.

7. Integrity and the Content of Right Action

In the foregoing cases it is accidental that performing an action would require a person to act against her integrity. If those actions have reasons in their favor and are done in the right way they *can* be done rightly and thus *can* be right action. This seems true of most actions. Yet Williams' complete account requires actions to exist that *in principle* cannot be done in the right way. Williams suggests integrity conditions whether certain actions can make up part of the content of right action. He hypothesizes that actions demanding the forfeit of one's integrity as such cannot be right, at all. To account for that relationship requires some actions to be ruled out as necessarily incorrect because they require the forfeit of one's integrity as such. They must be incorrect in virtue of that structure.

It seems clear how a demonstration of an action's structural incorrectness would go. The ordinary concept of right action seems to entail acting out of the intention to right action for its own sake, which entails acting on one's best judgment about what is right because it is right. This in turn entails acting out of the intention to act on one's best judgment about what is right because it is right, that is, with integrity. So action that cannot in principle be performed because

it is right as the outcome of the intention to act on one's best judgment about what is right cannot in principle count as right action. The structure of right action would rule it out as one of its possible contents, regardless of any considerations that might seem to count in the action's favor.

It further seems the necessity to which this "entails" refers must be a form of necessity contained in the action's own structure. Otherwise, circumstances could exist wherein the action could be the right thing to do. If the structure of right action is to rule out possible contents, it must do so in virtue of structure. The necessity in question must be constitutive. It must be that it does not make sense to perform an action with integrity. So for any action ϕ , ϕ must be or must entail *not* forming and trying to act on one's best judgment about what is right because it is right, *not* having integrity. Thus " ϕ is right" must be or must entail "It is right not to form or intend to act on your best judgment about what is right because it is right."

On the surface, this locution seems innocuous. One can think of several examples in which "It is right not to form or act on your best judgment" or even "it is right not to act with integrity" can express a very sensible recommendation. I first want to try to show that such innocuous recommendations *do* exist and are sometimes even correct. But I also want to suggest that other senses of those locutions cannot be. Differentiating these senses will be the first task of showing how and why the structure of right action as ordinarily conceived rules out as incorrect any actions that constitutively cannot be performed with integrity.

8. Action Constitutively Impossible for Integrity, Thus for Right Action

The conception of integrity formed above has two parts: forming one's best judgment about what is right and trying to act on one's best judgment about what is right.³⁸ Thus a person

³⁸ To avoid clutter, from here on I will omit "because it is right" as implied in every use of the phrase "act on one's best judgment about what is right," except where it becomes systematically important to mention it. There is no

can be advised to do something that could be called “not having integrity” either by being told “it is right not to form your best judgment about what is right” or “it is right not to try to act on your best judgment about what is right” or by being advised to do something that implies either of those statements or both. It seems each statement can have at least two different senses. On the one hand,

- (A) “It is right not to form your best judgment about what is right,” could mean either,
 - (A.1) It is right to perform a further-order project of not forming a lower-order best judgment about what is right, or
 - (A.2) It is right not to form a highest-order best judgment about what is right.

On the other,

- (B) “It is right not to try to act on your best judgment about what is right,” could mean either,
 - (B.1) It is right to perform a further-order project of not intending on a lower order to act on a lower-order best judgment about what is right, or
 - (B.2) It is right not to intend at the highest order or act on your highest-order best judgment about what is right.

I want to suggest now that (A.1) and (B.1) recommendations can describe perfectly correct actions. It is wholly constitutively possible to perform these actions with integrity, thus for those actions to be right. But (A.2) and (B.2) recommendations, I will argue, constitutively cannot be right. They constitutively cannot be performed with integrity and so cannot be right action.

8.1 (A.1) and (B.1) Recommendations

Under (A.1) recommendations, one is told one should actively control one’s lower-order judgments so one either (A.1.1) forms them less rigorously than one could, or (A.1.2) refrains from forming them about what is right, or (A.1.3) refrains from a lower-order judgement, at all.

For instance, someone recommends that one should not try so hard on the second order to

acting on one’s best judgment about what is right that is not acting on what is, one judges, right because it is right. One can acts in conformity with it on some other basis or arbitrarily, but this does not count as acting on one’s best judgment about what is right as such. I will point this out in cases wherein it becomes relevant.

determine the facts about what is right. One should control oneself from the third order in such a way as to allow one's first-order views about what is right to be at times less accurate than they could be. Yet one should take on this third-order project of controlling one's lower-order states because it is *right* to control those states in this way. Similarly, under (B.1) recommendations, one is told one should actively control one's lower-order intentions so as not to act on one's lower-order best judgments about what is right, either by actively (B.1.1) refraining from intending or acting on it, or by (B.1.2) guiding one's lower-order intentions or actions on some other basis, or by (B.1.3) refraining from lower-order intention, as such. In other words, one should control oneself from the third order in such a way as to intend on the second order to act on the first order according to some basis other than what appears from a lower-order perspective to be right. But, again, it is recommended that one take on this third-order project because it is right to control one's lower-order states in this way.

It seems easy to think of cases in which (A.1) or (B.1) recommendations can be true. Say one is nearing the end of life and, after a lifetime of thinking to the extent of one's abilities about what is right and trying to do it, one's doctor tells one to take it easy. By this time, she says, one will have developed as a person in such a way that what makes one happy will generally be right and, moreover, thinking as rigorously as one can about what is right and trying to act on that judgment is putting undue stress on one's heart. "You should just try to be happy," she says. The content of her recommendation is as follows. One should (A.1.1) actively take on a third-order project of forming one's second-order best judgment about what is right less rigorously than one could, and one should (B.1.2) actively take on a third-order project of guiding one's actions and lower-order intentions by what makes one happy, instead of what one's second-order

judgment presents as right. One should take on both these projects because they are the right things to do in these circumstances.

Both of these recommendations can describe right actions. They can do so because of our capacity for further-order best judgment. In the third order, one takes into one's consideration not just one's view as of the facts (first order) and not just one's view on one's view as of the facts (second order), but one's view on one's view as of the facts. One can think, for instance, about one's epistemic faculties, about the extent to which one is capable of coming to accurate views of the facts. One may come to see one's own ability to determine the facts about a subject matter as limited or as a non-ideal guide to right action and take another person's testimony, for instance that of this doctor, as one's guide as to what to do. It seems it may often be correct to leave off one's attempts to determine the facts about what is right for a time. Taking both the way the world presents itself to one *and* one's epistemic situation into consideration, one can form the best judgment that it would be right to control oneself for awhile in such a way as not to work to the best of one's ability to determine the facts about what is right and act on them. One can do this as the *outcome* of doing one's best to take both the world and one's mind into account in determining what it is right to do.

The doctor presents one with third-order advice: guide yourself epistemically and practically for a time by what makes you happy, rather than what appears to you as the facts about what is right. Taking that third-order advice into third-order consideration, one can judge as far as one can tell from the third-order that the doctor is correct. One can then take on the project of doing as she says *because* it is right because, so far as one can tell (from the third order), it *is* right. One in other words trusts the doctor: one takes it from the third order that she knows what she is talking about and one intends from the third order to act on her advice. This

would apply in any case in which one takes another's advice as to what is right when one cannot see it more directly for oneself. One does not refrain from making *any* best judgment. One makes the best judgment that someone else's word is decisive evidence against one's own lower-order views. One then tries to act on that third-order best judgment.

One can describe this structure systematically in the following way. If φ is right, then one constitutively must do φ because it is right. This constitutively requires one to form the higher-order best judgment (φ is right) and to intend (φ) because it is right. If φ is (A.1) "not to form a lower-order best judgment," then it is wholly possible for one to form a further-order best judgment (it is right not to form a lower-order best judgment about what is right) and intend that action because it is right. On the other hand, if φ is (B.1) "not to act on a lower-order best judgment," then it is wholly possible to form the further-order best judgment (it is right not to form a lower-order best judgment about what is right) and act on it because it is right.

(A.1) and (B.1) cases seem to cover most of the actual cases from human life in which someone recommends one not to try to do the right thing, not form one's best judgment about what is right, or not try to act on it. Yet such recommendations do not genuinely require one to act against one's integrity as such. They require one only to expand the ambit of the facts one takes into consideration in forming one's best judgment or change the order of the best judgment on which one acts. Only recommendations of type (A.2) and type (B.2) counsel us not to try *at all* to determine the right thing to do or to do it for its own sake. But those, I will now suggest, constitutively cannot make sense as right action.

8.2. Type (A.2) and (B.2) Recommendations

In contrast with the former two types, (A.2) and (B.2) recommendations do not tell one to actively impose further-order control over one's lower-order states according to a further-order best judgment about what is right. They tell one *not* to impose control over one's lower-order states according to one's best judgment about what is right, to impose control over one's lower-order states according to something *other* than one's best judgment about what is right, or not to impose control over one's lower-order states, at all. That is to say, they tell one not to form or not to act on one's highest-order best judgment about what is right. I will argue that it cannot be right not to be about the highest-order project of determining and doing what is right, whether by declining to form one's best judgment or by declining to intend or act on it. To do so is not to try to do the right thing. But one cannot do the right thing by not trying to do the right thing.

Type (A.2) recommendations tell one not to exert highest-order control over one's views about what is right, either as well as one can or at all. They recommend that one (A.2.1) should not form one's highest-order judgment about what is right as well as one could, or (A.2.2) should not form a highest-order judgment about what is right, or (A.2.3) should not form a highest-order judgment, at all. For instance, someone recommends that one leave aside the question about what is right entirely and think only about how to be successful. One should undertake to form *no* highest-order view about what is right but rather to form highest-order views about what will make one successful, merely as such. Similarly, type (B.2) recommendations say one (B.2.1) should not intend at the highest order or act on one's highest-order best judgment about what is right, or (B.2.2) should intend at the highest order or act on the basis of something other than one's highest-order best judgment about what is right, or (B.2.3) should intend at the highest order or act on some basis other than one's best judgment about what is right, or (B.2.3) should

not intend or act, at all. For instance, someone recommends that one should not on any level try to act on one's best judgment about what is right. One instead should intend at the highest-order to act on one's highest-order best judgment about what will make one successful, merely as such.

In a certain way, the reason why these actions cannot be performed with integrity and thus cannot be right action seems straightforward. Leaving aside human psychological capacities, one may always in principle take up a higher-order attitude toward the world, toward one's view of the world, toward one's view of one's view of the world, and so on. But at whatever point the order of one's attitudes ends is one's *own* perspective, the way one sees things. Not to form one's best judgment about what is right from this perspective (A.2) is to have *no* view as of the facts about what is right *as* the facts about what is right. Likewise, not to try from this perspective to act on one's best judgment about what is right (B.2) is not to try at all to act on what one sees as the facts about what is right. Yet intending or acting on *some* view as to the facts about what is right *as* the facts about what is right constitutes a necessary condition of acting on the *actual* facts about what is right because it is right. Thus it cannot be right not to form or not to act on some view as to the facts about what is right as the facts about what is right. Any putative right action that would constitutively require one not to do either of these things cannot be a possible content of right action. One cannot act rightly by failing to possess a condition necessary for acting rightly. Thus it is not and cannot be right action, at all. Yet this very abstract presentation seems much clearer when one considers individual cases.

8.3 Type (A.2) Recommendations

Consider the following case. Say one lives within a social order overtly or covertly waging a constant campaign to undermine the willingness and ability of its citizens to form their

best judgments about what is right. The regime in charge commands one constantly to passively accept what it says is right. Any sign of dissent is swiftly punished, so to think as best one can about what is right is a constant struggle fraught with danger. One day the idea comes to one that maybe one *should* just accept what they say. It would be so much easier just (A.2.1) not to pursue questions about what is right to the extent one is able or (A.2.2) to have no opinions about what is right at all or (A.2.3) to have no opinions, at all. One wonders whether the right thing to do would be just to let go and stop trying to work out as best one can what is right.

It seems clear no (A.2) recommendation can constitute right action once one considers the way the course of action would appear from one's own perspective. From the first-person perspective, one's best judgments are not just mental states. They present the world itself as best one can make it out, including one's own nature when addressed from the higher orders. To come to a best judgment about what is right is to see the world as containing certain facts about what is right and to see those putative facts *as* facts. To decline to form a best judgment about what is right, then, is not to come to such a view. One can do this by having a view as of the facts, but not seeing those facts *as* the facts because, for instance, one holds back from certain questions for the sake of an interest or compulsion, as Jim (1) and (2) did. One can also do it by coming to *no* view about what is right, whether that takes forms (A.2.2) or (A.2.3).

Yet it seems clear a person with integrity constitutively cannot fail to form a view as to the facts about what is right and to form it to the best of her ability. The only way she could fail to form that view would be if she already lacked the highest-order intention to do the right thing. Intending to do the right thing *is* to try one's best to determine what is right and do it, to form one's best judgment and try to act on it. But one cannot do the right thing because it is right without doing it out of one's intention to do the right thing. One cannot act out of an intention

one lacks. Thus (A.2) recommendations and recommendations that entail them are ruled out by right action as ordinarily conceived. They cannot describe its possible contents.

In systematic terms: intending (φ) because it is right entails having formed the highest-order best judgment (φ is right). One must have formed the highest-order best judgment (it is right not to form one's highest-order best judgment) in order to carry out the required course of action because it is right. But in forming that highest-order best judgment, one contravenes the requirements of the putative right action. Yet if one does *not* form that best judgment, one fails to perform the action, at all. One cannot fulfill the requirement because it makes no sense to fulfill the requirement. It thus makes no sense as right action, either.

8.4 Type (B.2) Cases

Consider a case similar to the one above. One lives in a repressive social order, but the regime governing it does not campaign to make its citizens believe things unquestioningly. Instead, it tries to force them to act or intend according to its will without regard for what the citizens themselves see as right. Through ever-present threats of psychological or physical violence, the regime tries to force its citizens to respond intentionally to its commands not because of any legitimate, reason-giving authority the regime might have, but merely because the regime, as such, has issued them. One subjects oneself to great and constant danger and comes into constant friction with one's society in trying to act on one's best judgment about what is right. One day one considers whether one should just give in. It would be far less difficult (B.2.1) not to try to do what, so far as one can tell, is right or (B.2.2) just to do whatever they say, not because they have the right to do so, but just to avoid the threat of punishment, and not

because avoiding the threat of pain would make it right, or even (B.2.3) just stop willing entirely. Could any of these be the right thing to do?

It seems clear no (B.2) recommendation can be true, considering the perspective of the person considering whether it is right. Again, from one's own perspective, one's highest-order best judgment about what is right presents the facts about what is right *as* the facts about what is right, so far as one is able to make them out. Because acting on a fact constitutively entails recognizing that fact, failing to intend or act on some such view as of the facts about what is right constitutively entails failing to intend or act on an *accurate* view of the facts about what is right *because* they are the facts. Failing to intend or act on one's highest-order best judgment about what is right thus constitutively entails that one fails to do what is right. One can do this either (B.2.1) by simply not intending to act on that highest-order best judgment, (B.2.3) by not intending at all, or (B.2.2) by intending or acting on some other basis.

Again, it seems a person with integrity constitutively cannot fail to intend to act on her highest-order best judgment about what is right. She could only fail to make that attempt if she already lacked the highest-order intention to do the right thing. One who has that highest-order intention already constitutively tries as best she can to determine what is right and act on it. Without that intention, she cannot do what is right because it is right. But again, one cannot act out of an intention one lacks. Thus, like (A.2) recommendations, (B.2) recommendations and recommendations that entail them cannot be parts of the contents of right action.

In systematic terms: intending (φ) because it is right entails that one intends on the highest order (to act on one's highest-order best judgment that (φ is right)) because φ is right. But it makes no sense to intend on the highest order (to act on one's highest-order best judgment that (it is right not to intend or act on one's highest-order best judgment about what is right))

because it is right. One contravenes the right action in intending it and fails it in not intending it. It makes no sense to fulfill it, thus it makes no sense as right action.

8.5 Structurally Incorrect Action

This structure seems to help complete the systematic account of the role Williams argued integrity must have in determining the content of right action. It was argued earlier that one does not count as acting rightly if one's actions contravene one's integrity in a particular situation. This structure suggests there are actions that one cannot act rightly in performing in any situation. They are not accidentally but structurally incorrect. Indeed, (A.2) and (B.2) recommendations cannot be right action even if the actions in question are otherwise correct or good ones. It could be imagined, say, that the governing regime campaigned to force people to accept *true* beliefs about what is right or to perform or intend actions that actually *are* right. The regime could be trying to induce people by psychological or physical force to "be good." But whether one believes something because it is easier to believe it rather than because, as far as one can see, it is true, or acts or intends in some way because it is easier not to act on one's best judgment about what is right, both entail not doing what one does because it is right. Even if the behavior is correct, one would do it because it is what, for instance, a coercive regime wants one to do or believe, or else not really to act or believe, at all.

Doing something because it is right requires doing it out of the intention to do what is right. One who intends to do what is right tries her best to determine what is right and do it, forms her best judgment as to what is right and tries to act on it. These are the first- and the third-person perspectives on the intention to do what is right, respectively. Actions requiring one to give up forming one's best judgment about what is right or trying to act on it thus

constitutively require one to give up the intention to do what is right. But to do the right thing, one must try to do the right thing. One cannot rightly by giving up trying. By the ordinary concept of right action, it makes no sense.

9. Conclusion

Williams argues from intuition that “integrity” must condition the content of right action. The fact that an action would require a person to act against her deepest commitments must play a role in determining what she should do qualitatively different than merely contributing or detracting from the value produced by her doing it. A person *cannot* be required to be untrue to everything she cares about in order to produce a more valuable world. Pure consequentialism cannot make sense of this qualitatively different role. Thus pure consequentialism must be mistaken.

How is this conditioning relationship between right action and integrity to be understood? On the one hand, it seems an adequate conception of integrity must satisfy at least two desiderata. It must mark important distinctions in the evaluation of action based on *how* persons’ commitments are formed, and it must make sense of the ordinary use of the word “integrity.” On the other hand, to explain the relationship between integrity and right action, that conception of integrity must bear a systematic relationship to right action as such.

I have argued for a conception of integrity satisfying the two desiderata by giving an account of the structure of right action. Ordinary language seems to suggest doing what is right constitutively requires doing it *because* it is right. Doing something because it is right seems, in turn, constitutively to require doing it out of the intention to do what is right because it is right. To intend to do what is right because it is right seems to be to try to the extent of one’s ability to

determine what is right and do it because it is right: that is, to form one's best judgment about what is right and intend to act on it. Right action is, in part, forming and trying to act on one's best judgment about what is right. Forming and trying to act on one's best judgment about what is right seems to be to have or act with integrity.

This conception of integrity seems clearly to satisfy the two desiderata. It makes a decisive difference both in the evaluation of action and in the ordinary attribution of integrity whether or not a person has formed her commitments as the outcome of her best judgment about what is right. Yet, when conjoined with the above account of the structure of right action, this conception also provides a systematic account of integrity's role in determining when a person acts rightly in any particular case. Whatever the preponderance of reasons in favor of a person's taking on an action φ , the person does not act rightly in φ -ing unless it is the content of her best judgment that φ is right and unless she undertakes that action on the basis of that fact, which she has thus acknowledged. She may perform the correct behavior, but she is not acting rightly. Action on best judgment is a necessary though not sufficient condition of correct performance.

Yet this only shows when otherwise correct actions do not count as having been genuinely carried out. It is a further question whether integrity can partially determine whether an action can constitute part of the content of right action in the way Williams thinks it does. I have argued it can. Since the ordinary concept of right action entails action with integrity, action that cannot in principle be performed with integrity cannot make sense as right action.

Action that cannot in principle be performed with integrity would be action the performance of which constitutively entails one is *not* acting with integrity. So for any putative right action to be in principle contrary to action with integrity, it must be or must entail that it is right not to have integrity: (A) not to form one's best judgment about what is right or (B) not to

try to act or intend on its basis. “It is right not to form your best judgment about what is right” and “It is right not to act or intend on the basis of your best judgment” can have two different families of senses, respectively. The first can mean either (A.1) that one should perform a further-order project of not forming a lower-order best judgment about what is right, or (A.2) that one should not form a highest-order best judgment about what is right, at all. The second can mean either (B.1) that one should perform a further-order project of acting or intending on the basis of a further-order best judgment about what is right, or (B.2) that one should not act or intend on the basis of one’s highest-order best judgment about what is right, at all.

“Not forming one’s best judgment about what is right” in the (A.1) sense and “not acting or intending on the basis of one’s best judgement about what is right” in the (B.1) constitutively *can* be performed by a person with integrity, out of the intention to do the right thing. One can form a further-order best judgment that takes, say, one’s second-order best judgment into account as a factor in determining what one should do, all things considered. In turn, one can form a higher-order intention to guide one’s lower-order intentions by that higher-order best judgment. One would do this, for instance, when confronted with testimony from a trusted individual that one’s view of the evidence is limited and so, for instance, one should act on her advice, even though one does not yet see why what she advises is correct. Integrity allows both for believing that one should do what does not seem right at the time *and* for intending to do that thing *because* it is right. One can take up an attitude toward oneself as an entity within the world with the epistemic limitations that implies and form a further-order best judgment in light of them. In that case, one does not literally stop forming a best judgment about what is right. One only takes one’s epistemic stance into account with all the other facts in forming one’s best judgment.

Yet neither the (A.2) nor the (B.2) senses can be performed by a person with integrity and thus neither can be right. One might be tempted to see it as right not to form a highest-order best judgment about what is right or not to try to act on it when it becomes difficult or even dangerous to continue to work out the facts about what is right for oneself or to live by one's own best judgment rather than, say, to accept the views an oppressive regime tries to compel one to take on by force or to obey it unquestioningly. But neither of these courses of action can be right. To perform either of them *because* it was right, one would have to perform it out of one's intention to do what was right. One thus would have to form the highest-order best judgment that the action was right and try to act on it because it was right. But in forming one's highest-order best judgment, one would contravene the conditions constitutive for performing the action. Yet if one did not undertake the action, one would fail to do it. Thus there is nothing it is like to perform the action because it is right. Thus it cannot make sense as right action.

A person with integrity forms her best judgment about what is right not necessarily because doing so has any value or is the right thing to do in its own right. She does so because she seeks to know what *is* right so she can do it *because* it is right. For such a person not to form her best judgment would be for her no longer to be about that project, for her to intentionally disregard her responsibilities as such. From her perspective, the question whether to continue forming her best judgment about what is right or to continue trying to act on it appears not as just the question whether to have certain mental states or not. From her perspective, the question whether to continue forming her best judgment or trying to act on it is the question whether to keep trying to do the right thing, that is, trying her best to determine what is right and do it. Yet one can only do what is right *out of* this attempt. It makes no sense to do the right thing by

giving it up. One may still fail to do the right thing if one tries. But one cannot act rightly in giving up trying. One cannot succeed in a project by forfeit.

So whether the best judgments or intentions of persons or personal integrity matter at all in their own right, one must in every case try one's best to determine what is right and live up to it. One must in every case think and act with integrity. One must do this not necessarily because it is right in itself. One must do it as part of one's attempt to do the right thing, to act in fidelity to what is right, whatever that is.

It may not be possible to say *a priori* what is right. But one can say *a priori* that one must try to do what is right. One must always try. One must try one's best.

BIBLIOGRAPHY

- Arpaly, Nomy. *Unprincipled Virtue*. Oxford University Press, 2003.
- Broome, John. "Normative Requirements." *Ratio* (1999) 12: 398-419.
- , -----, "Wide or Narrow Scope?" *Mind* (January 2007) 116: 359-70
- Davidson, Donald. "Incoherence and Irrationality" in *Problems of Rationality*. Oxford University Press, 2003: 189-98.
- Kolodny, Niko. "Why Be Rational?" *Mind* (July 2005) 144: 509-563.
- Korsgaard, Christine M. *The Sources of Normativity*. Cambridge University Press, 1996.
- Raz, Joseph. *From Normativity to Responsibility*. Oxford University Press, 2012.
- Scanlon, T.M. "Metaphysics and Morals." *Proceedings and Addresses of the American Philosophical Association* (November 2003) 77: 7-22.
- Twain, Mark. *The Adventures of Huckleberry Finn*, 1884.
- Wallace, R. Jay. "Normativity, Commitment, and Instrumental Reason". *Philosopher's Imprint* (December 2001) 1 (3): 1-26.
- Williams, Bernard. "Moral Luck." *Moral Luck*. Cambridge University Press, 1982.
- , -----, *Utilitarianism: For and Against*, with J.J.C. Smart, Cambridge University Press, 1973.