The Sage’s Psychology: Confucianism Naturalized

by

Daniel J. Stephens

Department of Philosophy
Duke University

Date:_______________________

Approved:

___________________________

David Wong, Supervisor

___________________________

Owen Flanagan

___________________________

Karen Neander

___________________________

Michael Ferejohn

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy in the Graduate School of Duke University

2015
ABSTRACT

The Sage's Psychology: Confucianism Naturalized

by

Daniel J. Stephens

Department of Philosophy
Duke University

Date: _______________________

Approved:

___________________________
David Wong, Supervisor

___________________________
Owen Flanagan

___________________________
Karen Neander

___________________________
Michael Ferejohn

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy in the Graduate School of Duke University

2015
Abstract

In this dissertation, I attempt to answer the question of how people can come to behave in accordance with their moral standards. To answer this question, I argue for and then apply a naturalistic approach to ethical philosophy that includes an attempt to construct both an empirically adequate account of human psychology and an account of moral cultivation that accords with that account of our psychology. I then present a part of that picture of human psychology, focused on what I call “impediments to virtue”, which are the elements of human psychology that make it difficult for us to behave in ways that consistently accord with our moral standards; this picture also serves to show why we need moral cultivation methods and helps to clarify what we need them to do for us. I then argue in favor of an interpretation of the Analects of Confucius on which it is primarily focused on discussions of a method of moral cultivation, and I lay out a detailed account of what that method is and how it works. Turning once again to literature in empirical psychology, I present an argument that we have good reason to think that the Confucian method of moral cultivation as presented in the Analects will be effective in the ways intended. I then discuss the relative strengths of the Confucian method over other methods of moral cultivation that exist in the philosophical literature, including Aristotle’s method of cultivating virtue in the Nichomachean Ethics, Mark Alfano’s factitious virtue theory, and biotechnological moral enhancement.
Contents

Abstract ......................................................................................................................... iv

Chapter 1: Moral Cultivation and Moral Theory ......................................................... 1
  1.1 Introduction .............................................................................................................. 1
    1.1.1 Aims .................................................................................................................. 1
    1.1.2 My Argumentative Strategies ......................................................................... 3
  1.2 The Umbrella Strategy .......................................................................................... 5
    1.2.1 PMPR and Ought Implies Can ....................................................................... 6
    1.2.2 Doris’ Abandoned Path: Specific Judgments of Reasonableness ................. 15
  1.3 The Disjunctive Strategy ...................................................................................... 23
    1.3.1 Easy Allies ...................................................................................................... 23
      1.3.1.1 Virtue Ethical Approaches ...................................................................... 23
      1.3.1.2 Peter Railton and Consequentialist Approaches .................................... 27
      1.3.1.3 Deontological Approaches .................................................................... 29
    1.3.2 A Harder Case: Constructivism .................................................................... 30
  1.4 Conclusion .......................................................................................................... 36

Chapter 2: Psychological Impediments to Virtue ....................................................... 38
  2.1 Introduction .......................................................................................................... 38
  2.2 Problematic Motives ........................................................................................... 41
    2.2.1 The General Picture ....................................................................................... 41
    2.2.2 Problematic Motives: A Few ......................................................................... 44
2.2.2.1 Self-interested Desires .................................................................44
2.2.2.2 Dominance ..................................................................................46
2.2.2.3 Punishment...................................................................................50
2.3 Problematic Cognitions.......................................................................53
  2.3.1 Cognitive Control and Ego Depletion.............................................54
  2.3.2 Biases..............................................................................................58
    2.3.2.1 Biases: A Few ...........................................................................59
      2.3.2.1.1 Self-Serving Biases.................................................................60
      2.3.2.1.2 In-group Favoritism and Outgroup Negativity.....................66
      2.3.2.1.3 Racial Bias, Gender Bias, and the Halo Effect.....................71
  2.4 The Problem of Situationism............................................................75

Chapter 3: Moral Cultivation in the Analects, Part 1: Is There a Normative Theory in the Analects? ..................................................................................................................83
  3.1 Introduction..........................................................................................83
  3.2 The Critical Arguments: That the Analects Does Not Present a Systematized Normative Theory ...............................................................................................85
    3.2.1 Normative Theorizing in the pre-Qin Dialectic..............................87
    3.2.2 A Basic Standard in the Analects?: On Dao, Li, and Yi..................93
    3.2.3 A Virtue Ethical Theory in the Analects? .......................................107
    3.2.4 A Role-ethical Theory in the Analects? .........................................117
    3.2.5 An Exemplarist Normative Theory in the Analects? ......................124
  3.3 If not a Normative Theory, Then What?.............................................132
Chapter 4: Moral Cultivation in the Analects, Part 2: The Method of Moral Cultivation in the Analects

4.1 Introduction

4.2 Rituals (li 禮) and Their Aims

4.3 Roles and “Rectifying Names”

4.4 Conclusion

Chapter 5: On the Effectiveness of the Analects’ Method of Moral Cultivation

5.1 Introduction

5.2 Support for the Analects’ Method

5.2.1 Attention – Claims (1) and (2)

5.2.2 Executive Control – Claims (3) and (4)

5.2.3 Emotional Insight – Claims (5) and (6)

5.2.4 Shaping the Emotions – Claims (7) and (8)

5.2.5 Roles and Role Attributions – Claims (9) and (10)

5.2.5.1 Rituals, Roles, and Social Contexts: Some Things to Note

5.2.6 Prescribed Environments – Claim (11)

5.3 The Confucian Approach in Relation to Other Approaches

5.3.1 Vs. Aristotle’s Method of Cultivating Virtue

5.3.2 Vs. Factitious Virtue

5.3.3 Vs. Biotechnological Moral Enhancement

References

Biography
Chapter 1: Moral Cultivation and Moral Theory

1.1 Introduction

1.1.1 Aims

The main point of this chapter is to provide arguments that bolster both the general claim that empirical psychology provides implications that ought to influence our normative theorizing and the specific claim that the results of a project of constructing and assessing empirically sound methods of moral cultivation are relevant to the content of our normative theories. Because anything that bolsters the specific claim obviously also bolsters the general claim, my arguments are intended to establish the correctness of the specific claim.

Part of my motive for making this argument stems from how I look at the types of questions that we need to ask and adequately answer in order to complete the ethical project, the project of using a philosophical approach to ethics to live better and/or more morally. As I see it, we need to ask “what” questions, i.e. the questions that are used to inquire about what one ought to do in a particular situation, what general standards or principles we ought to live up to, and so on. We also need to ask “why” questions, i.e. the questions of justification that ask why we should accept a particular standard or a particular moral judgment as correct. We can also, though, ask what I call “how” questions, i.e. questions of implementation that ask how we can come to live in accordance with the standards that we
have developed and justified through asking and answering our “what” and “why” questions.¹

In the history of Western ethics, the majority of philosophical attention has been paid to the “what” and “why” questions, with relatively little being given to the “how” questions. This is especially true of analytic philosophy since the early twentieth century, as psychology, education theory, and the like have largely departed from under the umbrella of philosophy; normative theorizing has largely continued as an enterprise focused on asking and answering the “what” and “why” questions of ethics. There appears to be a growing interest, though, in pursuing philosophical ethics in a way that pays attention to the “how” questions, which coincides with a growing interest in approaches to ethics that are informed by empirical psychology.

In the last quarter-century, to pick a few examples, Owen Flanagan, in *The Varieties of Moral Personality*, both advocated and enacted the project of pursuing a psychologically realistic ethics,² John Doris³ and Gilbert Harman⁴ each argued that situationist social psychology undermines virtue ethical theories by undermining the relevant conceptions of character traits upon which virtue ethical theories are built, Maria Merritt argued that a

¹ This categorization obviously is not an air-tight analysis of how language either is or needs to be used, as questions of justification could easily begin with the phrase “what justifies….” In fact, one might be able to do away with the word “why” altogether and replace it with the functionally equivalent “for what”. Similarly, “how” questions could be asked with the phrase “By what method…” or “In what way…” . Nonetheless, I find these to be useful, intuitive labels.
Humean virtue ethics can escape those criticisms, and Mark Alfano has recently lamented that “moral technology (his term for our answers to “how” questions) has gotten such short shrift, especially in recent analytic philosophy” and he has presented a corrective project. I agree that inquiry into the “how” questions and their best answers has the potential to be a quite fruitful project, both theoretically and practically, within the larger psychological realist program. As such, the arguments that I will present here should also be seen as lending yet another layer of support to the claim that this expansion of the focus of philosophical ethics is a philosophical move we ought to perform.

1.1.2 My Argumentative Strategies

There are two ways we might see the relationship between the project of normative theorizing and the project of developing accounts of empirically adequate cultivation practices as an answer to ethical “how” questions. One, we can call the “post-theorization” way; we do normative theory, find our preferred answer, and then we figure out how, given our psychology, we can cultivate compliance to the standards set out by our normative theory. The other, we can call the “mid-theorization” way; in the process of constructing our normative theory, our best accounts of moral cultivation imply that the content of our normative theory needs to have some shape rather than another.

---

4 See Mark Alfano, *Character as Moral Fiction* (New York: Cambridge University Press, 2013). Alfano picks up a project of psychological realism in ethics that was effectively launched in Flanagan’s *The Varieties of Moral Personality.*
It is generally accepted that accounts of moral cultivation can relate to normative theories in the downstream, post-theorization way. If this is the only way that they can relate, then it perhaps makes sense to see the project of developing methods of moral cultivation as separable from the project of normative theorizing, possibly to be passed off from philosophers to psychologists, life coaches, self-help authors, and so on.\(^7\) Whether it actually is the only way depends of course on whether we have sufficient reason to accept that accounts of moral cultivation are also related to normative theories in the upstream, mid-theorization way. I will focus here on establishing that we should indeed accept that accounts of moral cultivation are in fact related to normative theories in the upstream, mid-theorization way.\(^8\)

I have two main strategies for establishing this. One is what I call my “umbrella strategy”; this is an attempt to establish as a general metaethical principle that empirically adequate accounts of moral cultivation provide implications for normative theories that should affect those theories’ content. I will pursue this strategy in Section 1.2. I call the second strategy the “disjunctive strategy”; this is an attempt to establish that advocating any of a variety of individual approaches to normative ethics should also lead one to accept that, at some point, one ought to look at one’s normative theory in light of our empirically

---

\(^7\) Some people would argue that philosophy was always meant to be a transformative practice and that philosophers should thus care about cultivation even if it is an issue downstream of normative theory. Others would argue that such downstream discussions fall outside the proper purview of philosophical ethics. I am, however, less interested in what philosophers need to be doing to earn the title “philosopher” than I am in figuring out the best way to do ethics.

\(^8\) That’s not to say that, if we only assume the downstream relationship, we cannot still chase down something of both philosophical interest and relevance to the current discussion. For instance, I think that it could easily be argued that when we understand applied ethics as an attempt to generate prescriptions for certain particular actions in certain types of real situations, we can begin to see our psychology as just another part of our real situation, to be taken into consideration when making judgments in applied ethics or public policy. This, I believe, would lead us to say that prescriptions to morally self-cultivate and to help others to do so are some part of the answer to a large swath of applied ethics questions. That is not a result I would shake a stick at.
adequate accounts of moral cultivation and perform some set of suitable alterations to the normative theory as a result. I will pursue this strategy in section 1.3. The umbrella strategy should be sufficient, but finding support from other directions, as I hope to do with the disjunctive strategy, seems to me to be a reasonable way of shoring up my position. Doing so in this instance has the added benefit of giving a cursory account of some of the various directions that normative theorizing can take when we perform it with the appropriate stance toward empirically adequate moral cultivation practices.

1.2 The Umbrella Strategy

I will present two lines of argument in service of the umbrella strategy. In the first line of argument, presented in Section 1.2.1, I advance Owen Flanagan’s Principle of Minimal Psychological Realism (henceforth, “PMPR”), which tells us to “Make sure when constructing a moral theory or projecting a moral ideal that the character, decision processing, and behavior prescribed are possible, or perceived to be possible, for creatures like us.” I attempt to further support it by showing that it follows from the principle that ought implies can, or rather, a suitably limited analogue thereof, and I show that a complete, empirically adequate account of what is possible for creatures like us will include an understanding of our remarkable neuroplasticity and our ability to be shaped by our practices and other influences. This line of argument should show that since normative theories are limited by what is possible for us, and since an account of what is possible for us must contain an account of what results can be obtained by various practices, such accounts

9 Flanagan, Varieties, p. 32.
help to provide implications for what content our normative theories should ultimately contain.

The second umbrella strategy argument, which I present in section 1.2.2, is one that is open to people who, like John Doris, may not want to rest their psychological realism on “dicta in the neighborhood of psychological realism and ‘ought implies can’”\(^{10}\) because they think that such dicta “don’t do much to separate the philosophical goats from the philosophical sheep”.\(^{11}\) The argument I present resuscitates a possible argumentative strategy that Doris himself mentions and then discards. I argue that Doris is too quick to discard that argumentative strategy and that, when it is pursued properly, it gives us further support for the kind of psychological realism about ethics that I need to establish in order to establish the relevance of empirically adequate accounts of moral cultivation practices to normative theory.

1.2.1 PMPR and Ought Implies Can

The first thing that I would like to establish here is that accepting PMPR commits us to seeing adequate accounts of moral practice as providing implications for normative theories. Say we accept PMPR. We are then committed to saying that our normative theories should be limited by what is possible, or perceived to be possible, for creatures like us. In order to understand the limits that this places on our normative theories, we need to be clear about how to understand the possibility requirement, and the way that it is modified

---

\(^{10}\) Doris, *Lack of Character*, p.112.

\(^{11}\) Ibid.
by the “perceived to be possible” clause, as what is possible and what is perceived to be possible can sometimes differ. One way that they can come apart is when there are things that are perfectly possible for people, but are not perceived to be possible by them. For some people, any improvements in their character, behavior, station in life, etc. are perceived to be utterly impossible. I do not believe that these perceptions of possibility have any role to play in determining what we can ask of people, because this would create undue limits on what we can ask of people, limits that may far exceed the limits imposed by restricting our theories to asking of people only what is genuinely possible for them.

In other cases, however, cases in which what is perceived to be possible falls outside the scope of what is genuinely possible, the principle is less restrictive for the addition of the clause. We need to be clear about how and why we might want it to be less restrictive. If the restriction was loosened so much that what is genuinely psychologically possible became irrelevant to how we should limit our theories, being replaced by whatever we think is possible irrespective of the results of psychological science, it appears that it would undermine the point of imposing a limit of psychological possibility in the first place. A more minimal way that the “perceived to be possible” clause can be understood to loosen the restriction is by serving as a caveat that allows appropriate wiggle room between what our best theories tell us is possible and what, ultimately, is actually possible. This caveat makes sense, and expresses due epistemic modesty, if we acknowledge that our best current answers to the question of what is possible for us come from whatever the latest science tells us, which will constantly need updating, and this is how I believe we should generally understand the thrust of the principle. I would add to this understanding that one further
way that the clause can reasonably be seen to loosen the restriction is by allowing that one of
the things that we want our moral standards and ideals to do for us, aside from creating
action-guiding norms of conduct that help determine the demands we can make on one
another, is to serve to motivate us, and to this end, ideals like moral sainthood or full
compliance to our ethical ideals are valuable enough to remain a part of our moral view,
even if they are only perceived to be possible.12 Allowing for these motivating ideals is not,
it seems to me, incompatible with also restricting our normative theories to what our best
science says is possible for us when it comes to our various other purposes in constructing a
normative theory, so long as when we are engaged in constructing a normative view, we are
clear about the multiple goals of a normative view, and the places where opening the
restriction from genuine possibility to perception of possibility gives us something of value
without the cost of, say, placing demands on people that they cannot meet.

Having gotten suitably clear about how we should understand PMPR, I want to
argue that in accepting it, we are then also committed to saying that if an adequate and
complete account of what is possible for creatures like us must contain an adequate account
of how we can be shaped by various practices, then such accounts of how we can be shaped
do provide at least some implications that contribute to the determination of what limits
should be placed on our normative theories. In other words, PMPR ensures my position
insofar as I can establish that what is possible for creatures like us must be spelled out with
reference to the effects of moral cultivation practices.

12 Indeed, Flanagan told me in an email that this was his motivation for including the “perceived to be
possible” clause.
While I will make the case in far more detail in Chapter 5, I think it is fairly clear, prima facie, that certain interventions, many of them practice-based, have effects on our dispositions and abilities, and also that this can make possible for a person certain behavior patterns that were previously not possible. This is how learning and developing skills works, whether those skills are moral skills, language skills, social skills, computer skills, or what have you. It is now widely accepted in neuroscience that our brains are remarkably plastic. In one recent article, Alvaro Pascual-Leone et. al. argued that “plasticity is an intrinsic property of the nervous system retained throughout a lifespan and that it is not possible to understand normal psychological function or the manifestations or consequences of disease without invoking the concept of brain plasticity”\(^{13}\), further extending a view first advanced by William James in *The Principles of Psychology* in the late nineteenth century. With plasticity as such an established aspect of the psychological reality with which we are faced, it seems clear to me that what is possible for us is something that indeed varies with our available practices and how we engage them. So, that one should accept my position that accounts of moral cultivation provide implications for our normative theories does follow from one’s having accepted PMPR.

Since my position on the relevance of moral cultivation to normative theory follows from PMPR, defending PMPR provides one way to defend my position. So, how might one defend PMPR and establish that our normative theories cannot be plausibly developed independent of an account of what is possible for us? Well, first, as a limiting case, let us suppose that the correct ethical theory is one that it is utterly impossible for anyone to live

suppose that the ur-principle in ethics is something like “we ought never to use resources that others may benefit from using”. In this extreme case, nobody living and breathing would ever be doing the ethically correct thing, regardless of what they did. The theory, then, would never pick out any of a person’s action possibilities as the correct thing to do. We can reasonably object to such a theory on the ground that it would fail to fulfill its purpose if, as I think it is plausible to assume, the purpose of a normative theory is more or less to help us collectively determine how to use moral terms to discriminate between the behaviors that we want to accept and the behaviors that we do not want to accept.

This objection might be a little too quick, though. One could perhaps argue that the purpose of an ethical theory is to express the moral truth, and the moral truth might be that we are all consigned to this kind of moral failure. I would push back against this argument by saying that if the moral truth consigns us to moral failure, then it still seems that it would be incumbent upon us to develop some sort of sub-moral-yet-action-guiding standards to help us navigate our worlds. That is an important role that needs to be filled by something regardless of what the moral truth is, and whatever fills that role is what I am concerned with here, whether we call it “morality” (as I do), “sub-morality” (as one might), or something else.

As comforting as it is to have this viable fall-back position in a defense of PMPR, I do think that there is a way to press the point and say that whatever our best take on the moral reality happens to be, it will be one on which we have accepted that our normative theories are limited by what is possible for us. One potentially fruitful strategy for defending such a claim would be to find some sort of metaethical principle that everyone should accept.
and show that PMPR follows from it. One such principle that we could possibly defend and from which PMPR might follow is the principle that “ought” implies “can” (henceforth, “OIC”). So, does PMPR actually follow from OIC? It appears to, since PMPR says that we ought to do something only when we can do it given the limits placed on our behavior by our psychology. Clearly, “‘Ought’ implies ‘can-given-the-totality-of-all-limits-on-our-possible-behaviors’” implies “‘ought’ implies ‘can-given-the-specifically-psychological-limits-on-our-possible-behaviors’”. So, we can defend my position by defending PMPR, and we can defend PMPR by defending OIC. The relevant question then becomes whether we can defend OIC.

Why should we think that OIC is correct? Well, for starters, to many people, the maxim is intuitively correct. It simply strikes most people as self-contradictory to say, “X ought to Φ, but X cannot Φ”, and one could argue that this gives us some pretty important insight into the logical structure of our concepts “ought” and “can”. OIC is often simply taken as an axiom of deontic logic, and, as David Copp has said, it is “unclear how one could argue for it”\(^\text{14}\). This is perhaps why it is so often simply taken as a given throughout the literature.

If the maxim appears intuitively correct and acceptance of it is taken as a default, then our continued acceptance of it at least rests on whether or not we have any good reasons to reject it. We may have reason to reject it if it conflicts with other principles that we have reason to hold to as tightly as OIC. In the literature, OIC has been shown to

conflict with principles or claims with similar intuitive support. Bernard Williams, for instance, has argued that accepting the possibility of genuine moral dilemmas, i.e. cases in which somebody ought to do A and ought to do B, but cannot do both A and B, puts the principle of agglomeration, the principle that when somebody ought to do A and ought to do B, they ought to do A and B, into conflict with OIC.\(^{15}\) He chooses to resolve this conflict by abandoning the principle of agglomeration and preserving OIC. Others, when faced with the same conflict have chosen to accept both agglomeration and OIC, and give up idea that moral dilemmas are possible.\(^{16}\) Others, however, have accepted the existence of dilemmas and the truth of the agglomeration principle, choosing to deny OIC.\(^{17}\)

While we clearly need to give something up to resolve this conflict, the existence of this conflict is not necessarily the best reason to reject one of these principles. Someone may also, of course, reject any two or even all three on independent grounds. So, are there such independent grounds for rejecting OIC? I believe that there are (so we need not look too closely here at what speaks for or against the principle of agglomeration or the possibility of genuine moral dilemmas).

Walter Sinnott-Armstrong has convincingly argued that we can generate cases in which it is still intuitively correct to say that somebody ought to do A, even when that person cannot do A. For instance, if Adams promises at noon to meet Brown at 6:00, then it is the case at 1:00 that Adams ought to meet Brown at 6:00. If, at 5:00, Adams is more than an hour away from where he is to meet Brown, it is still makes sense to tell Adams at

---


5:00 that he ought to meet Brown at 6:00. Since Adams cannot meet Brown at 6:00, and yet it still makes sense to say that he ought to meet Brown at 6:00, Sinnott-Armstrong argues, “ought” does not imply “can”.18

If we accept Sinnott-Armstrong’s argument, does this mean that OIC then fails to provide us with a reason to reject an impossible-to-follow moral theory? Not exactly. While Sinnott-Armstrong’s argument does show that “ought” does not entail or presuppose “can” simpliciter, he argues that it leaves open the possibility that “ought” still conversationally implicates “can” when used for particular purposes. Using “ought” in an advising capacity, for example, still can be seen as implicating “can”, because even if it still makes sense in general to render a judgment at 5:00 that Adams ought to meet Brown at 6:00, it makes no sense to still advise Adams that this is what he ought to do. In using “ought” in blaming somebody, similarly, “ought” implicates “can” (or at least “could have”) because for Brown to tell Adams at 6:00 that Adams is blameworthy because he ought to be here still makes sense, unless Adams can reply that, for instance, he could not have met Brown at 6:00 because the roads were closed.19 Because the blaming “ought” is typically used retrospectively, when we use it we are implicating the claim that there was a time at which both “Adams ought to meet Brown” and “Adams can meet Brown” were true; this implication of “can” is pretty clearly one of the limits that help determine when the use of the blaming “ought” is apt.

If we agree with Sinnott-Armstrong that “ought”, in these contexts, conversationally implicates “can”, then we can ask the further question of whether this gives us sufficient

18 See Sinnott-Armstrong, *Moral Dilemmas*, sections 4.2.2 and 4.2.3.
19 See Ibid., section 4.2.4.
reason to accept PMPR. PMPR cannot rest on OIC, since OIC is not correct, but can it rest
on what remains in the wake of Sinnott-Armstrong’s argument, something we can call
ABOIC (advisory and blaming oughts imply can)? I think that it does, because generating
both advisory and blaming “ought”s is what a normative theory is supposed to do for us. A
normative theory needs to tell us what we ought to do prospectively, and it needs to tell us
what we ought to have done retrospectively. If “ought” still implicates “can” when used in
these circumstances, then any reasonable advice or blame that may be generated from a
normative theory must be advice to do something that we can do or blame for not doing
something that we, at some point, could have done.

If we cash out, in these terms, the view that to be an acceptable normative theory, a
theory must be action-guiding, we can say that an impossible-to-live-up-to standard, like the
one I discussed above or even a guiding-but-impossible-to-live-up-to character standard like
a saint or a sage, fails to guide action in this fuller sense of generating both prescriptions that
we can follow and apt judgments of blameworthiness when we fail to follow those
prescriptions. Either the theory would constantly generate judgments of blameworthiness
directed at people for failing to do what they could not do, or the theory would draw a
distinction between what is and is not blameworthy that is fundamentally separate from what
the theory says we ought to do. The former option violates the principle that the “ought” of
blameworthiness implicates “can”, and the latter option would force us to give up the
fundamental idea that we are blameworthy for and only for not doing the things that we
ought to do. Neither of these strikes me as an acceptable option, as neither would yield an
ethical theory that is action guiding in the sense of generating the advisory and blaming “ought”s that implicate “can”s. So, I think we should accept ABOIC.

Does ABOIC then support PMPR and thus support my position on the relevance of moral cultivation to moral theory? I believe it does. If ABOIC says that our normative theories should only generate “ought”s that advise us to do things we can do or blame us for things that we could have done, this clearly includes what we can or could have done given all the limits on what is possible for us, including psychological limits. So, even if we have two great reasons to reject OIC proper given its incompatibility with the conjunction of the possibility of moral dilemmas and the principle of agglomeration, and given our ability to generate counterexamples of the sort Sinnott-Armstrong mentions, we can accept these objections to OIC proper while retaining the core of what makes OIC seem intuitively correct, namely, the idea that when we use “ought” in an advising or blaming capacity, there does exist the implication that a person can do, or could have done, the things we are saying she ought to do or ought to have done. These implications do serve as reasons to reject ethical theories that are not limited by what is possible for human beings, even if we decide we must give up OIC as a strict matter of entailment between “ought” and “can”, so ABOIC does in fact support PMPR, which in turn supports my position.

1.2.2 Doris’ Abandoned Path: Specific Judgments of Reasonableness

The next umbrella strategy argument that I will give for my view, as I mentioned in the beginning of section 1.2, picks up a line of argument that John Doris mentions in his book, *Lack of Character*. In the book, Doris is specifically interested in showing that
situationist social psychology undermines virtue ethics, but in service of that project he also argues for the general methodological conclusion that “ethical reflection is well served by interaction with the human sciences”. While Doris is assessing various ways of supporting psychological realism in ethics, he rejects the idea that the support for the kind of scientific psychological realism that he attributes both to himself and to Flanagan can come from something like OIC. He offers that OIC “might be understood as a prohibition against unreasonable or unfair ethical demands, and be glossed as something like ‘ought implies can be reasonably expected to’”. He then argues that, when we attempt to understand a modified OIC understood in this manner, “application of the principle will in every case require substantive ethical argument about reasonableness. And this argument is not one the proponent of virtue ethics is bound to lose; if unattainable ideals of virtue inspire people to behave better, there is a pretty good case to be made for these ideals’ reasonableness.”

My contention here is essentially that, in so quickly dismissing the idea that the case for placing psychologically realistic limits on our normative theories can ultimately rest on an account of what we can reasonably expect people to do, Doris actually overlooks a viable strategy for arguing in favor of the kind of psychological realism that he, Flanagan, and I more or less agree in advocating. Before I lay out an argument along that strategic line, I first want to note why I am unconvinced by Doris’ reason for rejecting the strategy. One reason may simply be dialectical. Unlike Doris, I am not searching for something with

---

20 Doris, Lack of Character, p. 2.
21 See Doris, Lack of Character, p. 113.
22 See footnote 20 for Chapter 6 in Doris, Lack of Character p. 207. Doris here calls Flanagan’s project “a fine example of scientific psychological realism” despite the fact that, as Doris says, Flanagan’s “‘official’ statement of his position commits him only to minimal psychological realism”.
23 Doris, Lack of Character, p. 112.
24 Doris, Lack of Character, p. 112.
which to undermine virtue ethics, so a path that leads to a judgment that certain virtue
ethical ideals are reasonable is not something I specifically want to avoid; if I can argue for
psychological realism in ethics in a way that makes empirically adequate accounts of moral
cultivation practices relevant to normative theory, then I am happy, even if it turns out that
virtue ethics is still a viable approach to normative theory. Another reason that I am
unconvinced by Doris is that his argument here rests on an equivocation, and, without this
argument succeeding, I see no reason not to defend psychological realism on the basis of
judgments about what people can reasonably be expected to do.

The equivocation that Doris makes is between two importantly different sorts of
judgments about the reasonableness of an ethical view. Doris claims that psychological
realism might rest on a general prohibition against unreasonable or unfair ethical demands,
and says, rightly I think, that the scope of this prohibition would then need to be further
unpacked via some substantive argument about reasonableness. Doris thus appears to say
that there are limits on our normative theories that stem from judgments regarding whether
those theories place reasonable or unreasonable ethical demands on us. So, the basic
judgments about reasonableness at which such an argument would bottom out would be
judgments of the reasonableness of our various plays in the moral language game, such as
requiring, prohibiting, advising, and blaming (i.e. reasonable use of the “ought”s I discussed
above), the making of which constitutes the placing of ethical demands on ourselves and
others. In then saying that unattainable character ideals will be called reasonable, however,
Doris seems to be saying not that the ideals are useful for generating a reasonable set of
advisory and blaming “ought”’s, but rather that the ideals are reasonable as something like objects of striving, because taking them up has beneficial results.

This kind of equivocation is problematic precisely because objects of striving, generally, can easily be reasonable absent the sorts of possibility constraints that we would want to place on our use of advisory and blaming “ought”’s. It is reasonable for me to strive to play a particular piece of music flawlessly, or to run a mile in 3:30 flat, or to absolutely maximize the amount of good I do, because this striving inspires me work harder than I would otherwise. It is not reasonable for anyone, though, to demand any of these things from me in the way that we all demand that I do not kill or steal, nor is it reasonable for people to blame me if I fail to do these things. So, Doris uses the word “reasonable” to make an illicit move; because one kind of positive reasonableness judgment is applicable to virtue ethical ideals, he claims that an entirely different kind of negative reasonableness judgment must not be applicable to those ideals. That claim clearly does not follow. He then concludes that we must not be able to get to anything that looks like a psychological-realist limitation on our normative theories via such judgments of reasonableness, because in order to impose such a limitation, we would need to apply negative reasonableness judgments.

Because Doris’s argument is flawed in this way, I see no reason not to try to pursue the line of argument that he proposes but rejects. As such, I will now attempt to show how a general prohibition against unreasonable ethical demands can still serve to support the sort of psychological realism that I seek here to defend, even in light of the fact that making such a prohibition meaningful obviously requires us to construct some set of substantive ethical
arguments regarding the reasonableness of various ethical demands.

How can I show this? Well, when we make the individual substantive arguments about reasonableness that serve to give the prohibition its content, we are attempting to give some reasons that support some individual judgments (or sets of judgments) that the various prescriptions, censures, etc. that constitute our set of ethical demands are either reasonable or unreasonable. So, I would need to show A) that there is some line or lines of argument that can adequately support these reasonability judgments. Then, to secure my overall point that we can get from such arguments to the sort of psychological realism that I seek to defend, I would need to show B) that this psychological realism is one of the upshots of the line of argument that I have constructed.

To show A, we can first ask what goes into the reasonableness judgments in question. For starters, individual judgments regarding the reasonableness of our prescriptions, censures, etc. are obviously informed by some fundamental normative judgments; it is reasonable, initially, to judge that people ought to do, at base, whatever we think people ought to do, be it to limit suffering, act justly, or whatever else. In other words, we identify some set of ends as reasonable, and so we judge that advising people to aim toward those ends and blaming people for failing to so aim are both reasonable, and advising people not to aim at those ends and blaming people for aiming at them are both unreasonable.

These fundamental normative ends are clearly not the whole story, however, since we can easily recognize that the reasonableness of various prescriptions or censures often depends on identifying ends plus correctly identifying facts about the available means for fulfilling those ends. It is unreasonable of me, for instance, to advise people to randomly
stab strangers on the street in an attempt to get them to fulfill the end of decreasing suffering, because it is simply a fact that doing so does not decrease suffering. We can perhaps express the relevance of these facts about means and ends by expressing a principle like “to will the end is to will the means”.\textsuperscript{25} If we accept this, then we must accept that individual prescriptions, censures, and so on are reasonable if and only if they tell us that we ought to fulfill some reasonable end via one of the various true means for fulfilling that end. If it is reasonable to say that one ought to do X, and a necessary condition for doing X is to take a course of action that satisfies the set of true means for doing X, i.e. (M\(_1\) or M\(_2\) or M\(_3\) or …M\(_n\)), then it is reasonable to say that one ought to do (M\(_1\) or M\(_2\) or M\(_3\) or … M\(_n\)).

So, we do have an argument capable of generating reasons that adequately support various individual judgments that certain prescriptions or censures, and thus certain sets of moral demands, are reasonable or unreasonable; I have thus shown A. Can I now show B? Does recognizing that we have an argument capable of supporting the relevant reasonability judgments give us a way of supporting the sort of psychological realism in ethics that I seek to defend, \textit{pace} Doris?

I believe it does. The argument so far makes it clear that, in general, facts about the world can make certain prescriptions, censures, etc. unreasonable, in ways that support the general prohibition Doris considers, either because in making these prescriptions, censures, etc., we will an end for which there is no means, or because in doing so we will a means that will not actually serve the end. This seems to apply for ordinary physical facts, because it is unreasonable to censure somebody for say, not saving a child in a situation in which doing

so would require an entirely superhuman feat of strength, or to advise somebody to, say, contribute to ending world hunger by salting local fields. It also seems reasonable on similar grounds to censure somebody for, say, feeding their children cyanide, given the combination of the reasonable end of giving their children proper nutrition and the fact that cyanide is not, as it were, part of a balanced breakfast.

In addition to these ordinary physical facts, judgments of the reasonableness of various prescriptions and censures can also depend on social facts about means (the end of respecting one’s grandmother is poorly served by calling her certain names), and, for instance, facts about means dependent upon how substances affect people (while it is generally reasonable to demand that I drive an injured friend to the hospital, it is always unreasonable to demand that I do so if I am heavily medicated, because doing so is no longer a means of fulfilling the end of protecting my friend’s wellbeing). If we accept this, then I think that we must also accept that psychological facts bear on these judgments in a similar way; it seems to follow naturally, and I can think of no special reason to exclude facts about psychological means from this sort of means-ends argument. If it is unreasonable to demand that I do something for which I have no available physical means, it is similarly unreasonable to demand that I do something for which I have no available psychological means, especially if one considers our present psychologies to be merely another part of the fact-pattern of the world. Similarly, if it is reasonable to demand that I fulfill some end, it must be because there is a potential means for me to do so given facts about my psychology. So, the sorts of limits that psychological realists would like to impose on our normative theories are the sorts of things for which we can argue by way of a prohibition on placing
unreasonable ethical demands on people, even when that prohibition must be unpacked in terms of individual judgments regarding the reasonableness of our various potential moves in the moral language game.

Does this, in turn, support my main contention that the results of a project of constructing and assessing empirically sound methods of moral cultivation are relevant to the content of our normative theories? I believe so, because to spell out these cultivation methods is simply to add detail to our picture of the available means we have for fulfilling our proposed ends, much like an understanding of farming techniques adds detail to our picture of how we might be able to eliminate certain forms of human suffering. Understanding these cultivation methods, and the necessary background of psychological facts that allow for various cultivation methods to be possible, effective, etc., helps us to understand, via the same kind of reasoning about means and ends that I have discussed throughout this section, when various prescriptions, censures, etc. are reasonable, and when they are unreasonable. So, if we accept a general prohibition against unreasonable ethical demands, and we give content to this prohibition via the substantive arguments about reasonableness that I have discussed here, one upshot is indeed that understanding, building, and assessing empirically sound methods of moral cultivation is a relevant part of normative theorizing. As such, I believe that both of my umbrella strategy arguments work, and so I shall now move on to discuss the arguments that comprise my disjunctive strategy.
1.3 The Disjunctive Strategy

While my umbrella strategy involved trying to show that, as a general metaethical matter, we ought to be committed to the claim that a proper part of normative theorizing involves constructing and assessing psychologically valid accounts of moral cultivation and then seeing what implications these accounts provide for our normative theories, my disjunctive strategy involves canvassing several individual approaches to ethics and showing that, by their own lights, they ought to be committed to the same sort of claim. This “bottom up” disjunctive strategy is, I believe, an important supplement to the “top down” umbrella strategy, in part because it provides further argumentative resources that should help to further convince specific groups of philosophers to view moral cultivation practices as relevant to moral theory, and in part because making these individual bottom up arguments serves to at least outline what it will mean for proponents of various approaches to ethics to take moral practice seriously in the manner that I am proposing. I divide this disjunctive strategy argument into two parts. I first discuss approaches to normative ethics that I think are easy allies in this discussion. In section 1.3.1.1, I discuss virtue ethics; in section 1.3.1.2, I discuss an argument advanced by Peter Railton that speaks in favor of consequentialists agreeing with my point, and in section 1.3.1.3, I discuss deontology. I then discuss, in section 1.3.2, the more difficult case of a constructivist approach to ethics.

1.3.1 Easy Allies

1.3.1.1 Virtue Ethical Approaches
It should not be too difficult to convince virtue ethicists that accounts of moral cultivation are an important part of a philosophical approach to ethics, since one thing that often draws philosophers to virtue ethics as an approach to normative theory in the first place is that virtue ethics, in focusing on what sorts of character traits one ought to have, makes conspicuous the need for development. Anna-Marie S. Christenson makes this perfectly explicit, saying that “Most forms of virtue ethics are characterized by two attractive features. The first is that proponents of virtue ethics acknowledge the need to describe how moral agents acquire or develop the traits and abilities necessary to become morally able agents.”

It is worth asking, however, whether this need to describe how people can develop the traits in question is a downstream, post-theorization need that results from the fact that having a certain kind of character requires developing that character, or whether it is also an upstream, mid-theorization need that means that the eventual content of a virtue ethical theory results in part from implications of an empirically adequate account of our available moral cultivation methods. I, of course, agree with the former view and want to argue in favor of the latter, which has not been a conspicuous possibility, even among virtue theorists.

That it has not been a conspicuous possibility is clear when looking at, for instance, how Thomas Hurka presents the landscape of the discussion about virtue in *Virtue, Vice, Value*.

---

26 This is often taken as a truism in the literature. This is made perfectly explicit in Anna Marie S. Christensen, "Getting it Right in Ethical Experience: John McDowell and Virtue Ethics," *Journal of Value Inquiry* 43 (2009), p. 493.
“One view defines virtue as a disposition to produce what is otherwise good or to do what is otherwise right, thereby giving it only a derivative and instrumental significance. Virtue may be crucial practically, if inculcating it is the best means of ensuring that people fulfill their moral responsibilities. But theoretically it has no intrinsic importance. The contrary view makes virtue the central property in a distinctive moral theory called virtue ethics, which is proposed as a fundamental alternative to consequentialism and deontology. Far from defining virtue as a means to goodness or rightness, this view treats it as primary, so that what is right and even what is good are identified by some relation to virtue.”

The first view is exemplified nicely by, for instance, the theory that Julia Driver presents in her book *Uneasy Virtue*, wherein she argues that the character traits that we identify as virtues are the ones that systematically produce good consequences in the world, and the second view is nicely exemplified by, for instance, Michael Slote’s theory in *From Morality to Virtue*, in which he argues that evaluations made in agent-based terms are privileged over evaluations made in action-based terms.

Hurka attempts to find a third way, arguing that virtue and vice are intrinsically good and evil respectively, but in a way that still allows such virtue judgments to fit within a broader consequentialist or deontological theory. My claim here, and in this chapter generally, is that there is a fourth way to view virtue; whether or not the normative standard we want is a characterological standard, an account of virtue as the character traits that lead one to act in a morally correct manner does have the kind of theoretical importance that Hurka mentions.

---

30 Hurka, *Virtue, Vice, and Value*, pp. 3-4.
While, again, virtue ethicists appear largely convinced that the project of assembling and assessing cultivation methods is important, and should thus be happy to see the kind of account that I present in subsequent chapters of this dissertation, it is still worth it to ask whether virtue ethicists qua ethicists have reason to be on board with my “fourth way” of relating our possible characters and how to cultivate them to the content of a virtue ethical theory. I think they do because of the following argument, if not also for other possible arguments that I will avoid making here for reasons of space.

Any virtue ethical account, to be worth its ink, is going to present a coherent set of character traits that people can actually embody, or at least can coherently aspire to embody and work toward embodying. If we want to adequately detail the process of cultivating those character traits, then we must draw from an accurate account of human psychology when we detail that process. Given many of the complexities of our psychology, and the resultant complexities that will be contained in an accurate account of the cultivation process, I think that an argument can be made that there are certain “process oriented” virtues that a person must have and develop in order to develop the more core virtues. For instance, if virtues like generosity, courage and the like take work to develop, then developing other, process oriented, enabling virtues like diligence, perseverance, a work ethic, etc. must be a part of the virtue theory. If those virtues must be a part of the theory, and those virtues must be a part of the theory because of the role they play in an empirically adequate method for cultivating virtue that is built in accordance with our best psychological picture of ourselves, then an empirically adequate method of moral cultivation, and thus also
empirical psychology generally, provides implications for the overall shape of the virtue ethical theory.

1.3.1.2 Peter Railton and Consequentialist Approaches

I believe that consequentialists also have reason, by their own lights, to see the content of their normative prescriptions as being influenced by psychological facts about human beings and what sorts of cultivation methods are available for altering our behavioral dispositions. There is already a strong argument in the literature that I think, mutatis mutandis, can be put to use in establishing this claim. That argument is found in Peter Railton’s “Alienation, Consequentialism, and the Demands of Morality”.

According to Railton’s argument, people attempting to accord with an objective consequentialism will have reason, from the moral standpoint engendered by that objective consequentialism, to take up an approach to choosing actions that does not always enact a consequentialist decision procedure, such as in the case of one choosing to act immediately during emergency situations rather than attempting to deliberate about which action would lead to the best consequences.31 The disposition to follow the quicker, non-subjectively-consequentialist procedure is recommended in part because, given the psychological realities of human decision-making, adopting this disposition leads, on the whole, to better consequences. In this case and others, Railton argues, “The objective act-consequentialist

would...recommend cultivating dispositions that will sometimes lead him to violate his own criterion of right action.”

Railton extends and illustrates this claim with the example of Juan and Linda. Juan and Linda have a long-distance marriage and fly to see one another every other week. One week, Linda is particularly depressed, and so Juan flies to see her, using money for the ticket that otherwise could have been sent to OXFAM and put to use producing better overall consequences. Railton argues that doing so still makes sense by consequentialist lights, because

“If Juan had had a character that would have led him to perform the better act (or made him more inclined to do so), he would have had to have been less devoted to Linda. Given the ways Juan can affect the world, it may be that if he were less devoted to Linda, his overall contribution to human well-being would be less in the end.”

Here, again, facts about the psychological reality of Juan’s possible characters and the behavioral outputs that Juan would regularly exhibit given those characters leads the objective consequentialist, by her own lights, to say that Juan ought to cultivate a certain kind of character that, sometimes, will lead Juan not to maximize good consequences.

This argument supports my main claim in this paper because it serves to establish that the overall shape of the content of (at least) act-consequentialist prescriptions is subject to possibility constraints that are in place as a result of facts about human psychology. With this established, it takes very little to move to the further claim that part of what informs the shape of those possibility constraints is what kinds of moral/psychological cultivation methods we have at our disposal. If Juan, for instance, had been able to cultivate a greater

32 Ibid.
33 Ibid., p. 159.
capacity to cheer Linda up, such that he would have been sufficiently consoling over the phone during that one week to keep Linda happy and to keep himself from being bogged down by guilt while still embodying the capacity for attachments that lead him to produce the good consequences in the rest of his life, then the individual consequentialist prescription that he send the money to OXFAM rather than use it to fly to Linda that week would have still obtained. Also, the overall consequentialist prescription about what kind of character Juan should have would include the judgment that he ought to have the more effusively consoling character, and that he should perform the sorts of cultivation practices that would allow him to have that kind of character. So, I believe that consequentialists, by their own lights, have good reasons to agree with my claim about the relevance of empirically adequate moral cultivation practices to the shape of their normative standards.

1.3.1.3 Deontological Approaches

I have already discussed, in section 1.2.2, how acceptance of the principle that to will the end is to will the means should lead one to the conclusion that I am trying to make in this chapter, so I need not belabor the point here too much further. I wish only to point out here that, insofar as deontological approaches to normative theory tell us what our duties are and tell us that we ought to perform our duties, they also tell us that we have subsidiary duties to perform whatever actions qualify as the means to perform our duties. This means that, given human psychological realities, the principles that deontology tasks us with following will very often also tell us to perform various cultivation practices that will allow us to follow the principles. This means that the total shape of the vast majority of
deontological normative views should also include principles the content of which is informed by our empirically adequate accounts of available moral cultivation methods. This also means that the more basic principles of a deontological view will be influenced by what our empirically adequate cultivation methods make possible for us, because to say that one ought to follow a principle for which there is no means, or, more on point, to say that somebody ought to follow number of seemingly compatible principles for which there would be conflicting means as a result of the messiness of our psychologies, would lead to an incoherence in the principles that one advocates that few deontologists would tolerate.

1.3.2 A Harder Case: Constructivism

While it should be rather clear that people who advocate virtue ethical, consequentialist, and deontological approaches to normative ethics should see accounts of moral cultivation as affecting the final shape of their normative theories, it takes a little more argument to support the claim that people who advocate constructivism in ethics should also accept my main claim of this chapter. I will present, in this final argumentative section, some reasons to think that they should.

To do so, because “constructivism” connotes a number of theories and positions, I first need to be clear what the theories or positions are about which I am arguing. Metaethical constructivism is the view that whatever correct answers there may be to normative questions, they will be determined by some definable decision procedure performed from some definable practical standpoint rather than by some set of discoverable normative facts out in the world; the normative theories that such a procedure yields are
what we can call constructivist normative theories. My argument here is essentially that if we start reasoning about ethics with the set of commitments about what can count as a viable normative theory that lead people to accept constructivist normative theories, then we ought to end up at a psychologically realistic normative theory, the content of which is influenced by our best available accounts of moral cultivation. The claim that I want to establish in this section can be understood as the claim that if one builds and justifies one’s normative theory via a constructivist decision procedure, then the content of one’s resultant normative theory should be shaped in part by implications provided by our best empirically-informed accounts of moral cultivation, just as in the case of the other normative views that I have discussed.

There are, of course, many ways of building a constructivist normative theory. In the interest of space, I will here focus primarily on the set of constructivist theories that are largely Kantian in approach, as the available conventionalist, Humean, and Aristotelian forms of constructivism appear to be a good deal more conducive to an approach that incorporates results from empirical psychology than the Kantian approach is; to convince the Kantian constructivists would be, then, to make more philosophical hay. In building a Kantian constructivist normative theory, theorists generally attempt to satisfy the following set of conditions to the greatest degree possible and with minimal trade-offs, or, at least, the most desirable balance of trade-offs, and it is this set of conditions that I am interested in here.

---

34 While it is beyond the scope of this chapter to spell this out, it will probably be clear to the reader that, mutatis mutandis, much of what I will say about Kantian constructivism could also provide support for the idea that proponents of these other forms of constructivism should accept the main claim of this chapter.

First, Kantian constructivists task themselves with finding a decision procedure that can be used to pick out basic norms. They also try to characterize that decision procedure in a way that refrains, as much as possible, from importing baseline normative content into the decision procedure; smuggling in realist normative assumptions, after all, undermines the point of identifying such a decision procedure as an alternative to accepting a realism that problematically countenances human-independent normative facts. They furthermore attempt to show that the basic norms that the procedure identifies support the largest possible chunk of our pre-theoretical normative judgments, hopefully leading to as easy a reflective equilibrium as we can find between the normative judgments that we accept and the basic principles identified by the procedure. This attempt to identify a decision procedure that picks out basic norms that fall into equilibrium with the normative judgments we accept, but without proliferating realist assumptions, leads the decision procedure to be specified as one in which agents, characterized in some minimal way so as to not assume too much normative content, choose these basic norms via some rational deliberation process.

The adoption of this basic strategy, and the related general aim of optimally navigating these constraints, means that building the various possible constructivist normative theories rests ultimately on finding different characterizations of agency, i.e. characterizations of the contractors who will rationally choose our norms, that lead to the contracting agents choosing basic norms that support our pre-theoretical judgments, but without purchasing that result by way of building disputable normative content into the characterization of agency. If we say that our contracting agents, for instance, prefer a

---

particular kind of life, have a particular conception of the good, are psychologically disposed
to value certain things, or so on, then the normative content of the output will simply be a
reflection of the normative content in the input, and we will have failed to ground our
normative judgments on an objective procedure.

This need to find a minimal conception of agency essentially bars empirical
psychology from bearing on a conception of the contracting agents, lest the theory collapse
into a sort of teleology. If empirical psychology is barred from bearing on the conception of
the contracting agents, then, of course, accounts of moral cultivation methods built in light
of empirical psychology are also barred from bearing on the conception of the contracting
agents. So, if empirically adequate accounts of moral cultivation methods do provide
implications for the content of these constructivist normative theories, it cannot be because
they help to inform our conception of the norm-choosing agents. So, how might they still
influence the content of these constructivist normative theories?

To answer this, we need to look not only at who is doing the choosing of norms, i.e.,
what the best conception of our contracting agents is, but also for whom the contracting
rational agents are choosing norms. Here, I see two distinct possibilities. We can either
conceive of the decision procedure as one in which rational agents are choosing norms for
their rational selves, who, being rational, we can think of as consistent and reliable agents, or,
we can conceive of the decision procedure as one in which rational agents are choosing
norms for actual human persons, whom we now have far more reason to see as inconsistent,
unreliable agents than to see as consistent and reliable agents. My claim here is that the latter
conception of the decision procedure is the most defensible by constructivist lights, and that,
when we build a constructivist normative theory according to that latter conception, the content of the resulting theory will be influenced by empirically adequate accounts of moral cultivation methods.

The first thing to point out in supporting my claim is that neither of these conceptions of the decision procedure presents a clear default. One might think that the former conception does, because our contractors are supposed to be rational agents choosing norms for themselves to live by, but the rationality we attribute to them here does not extend from them qua choosers to the them qua actors unless we make a number of assumptions about action theory and the psychological relationship between a capacity for rational deliberation and a capacity for consistently rational action. Both conceptions, in other words, rest on some positive views of the agents as choosers of the norms, the agents as followers of the norms, and how these two relate, and there is no reason to simply think that the placing of rationality constraints on agents as choosers implies a similar constraint on agents as actors unless we simply and unjustifiably collapse the distinction between choosing and acting. Since the choice between a conception of the decision procedure as one in which rational contractors choose norms for rational actors and a conception of the decision procedure as one in which rational contractors choose norms for human actors with all of the irrationality, inconsistency, and unreliability that we actually display must be decided by some positive argument, I will now present an argument that I believe should lead us to choose the latter conception.36

---

36 In addition to the argument that I provide in the following paragraph, I would also point the reader to a similar idea in Peter Railton’s paper “Moral Realism”, in which he discusses a maximally rational Lonnie-Plus that chooses courses of action for far less rational Lonnie. While Railton is not advancing a constructivist view,
Since Kantian constructivists will be committed to the idea that the contracting procedure will be one that is performed by rational agents, my argument starts by imagining these maximally rational deliberators discussing what norms to choose. In that situation, I think it is clear that rational deliberators, given a choice to either have access to the best available psychological picture of the agents who will be following the norms or to forgo that access, will choose to have that access. It then also seems clear to me that, given access to the best available psychological picture of human agents, which overwhelmingly shows us to be unreliable and inconsistent absent some interventions which will most likely include various cultivation practices, moral technologies, etc., the rational deliberators will choose some basic principle or principles and a norm or norms that prescribe some sort of moral/psychological cultivation practices. To choose some set of norms for creatures like us, and not also include a prescription to cultivate, would be to irrationally choose to exist in a world that is morally sub-optimal. This also means, I believe, that the content of principles that are chosen will also be informed by this need to cultivate, because the prescription to cultivate must not conflict with the basic principles, since they must exist as a rationally coherent whole in order to be chosen by maximally rational contractors. In other words, the kingdom of ends that we will cannot be one in which the laws cohere but nobody follows them, and in order to avoid this, given that people are not rational actors, we need to will that everyone engages in the psychology-shaping cultivation practices, and willing this influences the shape of the moral principles that we choose.

I think his discussion can be seen as lending support to the claim I am making here. See Peter Railton, "Moral Realism," The Philosophical Review 95, no. 2 (1986): 163-207. I thank David Wong for bringing this connection to my attention.
Note that this way of putting psychological facts to use in a constructivist procedure, while it does affect the eventual normative output, does not import any additional normative content into the notion of rational agency used to characterize the agents engaging in the deliberation over and choice of basic moral principles. It therefore does not run afoul of any of the basic goals or constraints that constructivists already largely accept qua constructivists. Whether a constructivism that proceeds this way will yield principles that actually support and fall into the right kind of equilibrium with our pre-theoretical normative judgments is, of course, an open question, but it is one that requires substantive normative debate that is outside the scope of this chapter. The important thing to note is that, insofar as one is a constructivist as I have outlined, one has sufficient reason to accept this psychologically realistic way of being a constructivist, and this seems to me to be so regardless of either how one decides to conceive of the rational contractors or decision procedure to get the normative outputs right, or how one sees constructivism as providing an answer to general meta-ethical questions about normative truth.\(^\text{37}\) Constructivists of various stripes, then, should largely agree with me.

1.4 Conclusion

If the arguments that I have presented in this chapter work as well as I believe that they do, then I have sufficiently established the philosophical importance of empirically adequate accounts of moral cultivation methods, both as a general metaethical claim (in my

\(^{37}\) Although, again, I stress that, within the scope of this paper, this need only be thought of as a hunch as it pertains to conventionalist, Humean, and Aristotelian constructivisms, although I take it to be a quite plausible hunch.
umbrella strategy arguments) and as a set of specific claims that should be accepted by various groups of normative theorists (in my disjunctive strategy arguments). Over the next four chapters, I will attempt to show that on the best reading of the *Analects of Confucius*, what that text gives us is a method of moral cultivation that we have good reason to accept as empirically adequate, giving us not only a method of moral cultivation that can help us to be more moral by most of our own lights, but also a plausible way of detailing our answer to the question of what is possible for creatures like us that provides justified limits on the content of our normative theories.
Chapter 2: Psychological Impediments to Virtue

2.1 Introduction

If we were more fortunate as a species, things would have been different. We would have existed in a world in which every one of our needs could be met effortlessly, in which we could not individually benefit by making others worse off, and in which we were utterly incapable of harming each other regardless of what we did. In such a world, morality would have been a completely useless notion.¹

The world that we do live in, of course, is one in which it requires a good deal of effort to meet our needs, our lives are saturated with possibilities for zero-sum and negative-sum exchanges, and we can each be easily harmed by another person in an enormous variety of ways.² In such a world, morality is useful because it helps us to secure our access to what we want and need and serves to safeguard us from harms. Though we have had the bad fortune of coming to exist in a world in which morality is so indispensable to our wellbeing, if we had been fortunate in other ways, this may not have been such a problem. We could have been given a psychology simple enough that we could simply decide to act in

¹ Cf. David Hume, *A Treatise on Human Nature*, ed. David Fate Norton and Mary J. Norton (New York: Oxford University Press, 2000), Book III, Section ii. While Hume’s point is of course only about justice, I think that in the conditions that I have given, we could generalize the spirit of his point to all of morality.

accordance with some morality, and this would be all that it took to ensure compliance for the rest of our lives.

We were not fortunate in this way, either, though. In actual fact, we are the products of evolution, and as such, we have a psychology that has been cobbled together in response to a variety of selection pressures. This has left us with a conflicting mass of abilities, drives, emotions, cognitive biases, heuristics, and so on, existing in such a way that our consistent accordance with a justified morality is fairly difficult, or at least quite rare; in other words, we are just not hardwired to be virtuous.\(^3\)

How, then, are we wired? If, by this question, we mean to ask for a complete picture of our psychology, then this question is probably, at present, impossible to answer satisfactorily. We can, though, replace it with a more manageable question, one that serves the goals of this dissertation project: Which elements of our cobbled-together psychology tend to impede us if we seek to become virtuous? In this chapter, I will attempt to give a concise but useful answer to that question. Answering this question here serves two functions. First, it supplements my argument from Chapter 1 for the importance of answering the “how” questions of morality by giving detail to our understanding of why exactly we need a method of moral cultivation and what specifically we need such a method to do for us. Second, by giving us a picture of what problems we require a method of moral

---

\(^3\) I should note that I do not, in this dissertation, endorse any particular answer to the “what” questions of normative ethics. When I discuss our impediments to virtue, by which I mean our difficulties in acting in a morally consistent manner, I am depending on a sort of minimal moral consensus or minimal moral common-sense-ism. Where such consensus does not exist, I am happy to be a pluralist and say that elements of our psychology often get in the way in a person’s attempt to act in accordance with his or her conception of morality, and that different elements of our psychology will present obstacles of varying difficulty to different people depending on their varying normative views.
cultivation to solve for us, it will help me to assess in later chapters what I see as the Confucian method.

The rest of this chapter is divided into three sections. In 2.2, I discuss drives that we have that sometimes incline us toward immoral actions. The general picture presented here will be that we humans are goal-oriented creatures, saddled with a motivational system that adapted to get us the things that we needed to successfully survive and reproduce, and there are several ways in which this particular motivational system can lead us in a counter-moral direction. In 2.3, I will discuss some elements of how we cognize that also make consistently moral behavior difficult. The upshot of this section is that the processes by which we cognize have also adapted in ways that help get us the things we need, so as we cognize about our material and social worlds, many of our cognitive processes put a thumb on our epistemic scales, so to speak, making us more persistent in the competition for resources by making us more likely to believe certain things over others. In 2.4, I will discuss situationist social psychology. While situationism has largely been presented as undermining character ethics by showing that they presuppose something false about character traits, the upshot here will be that our susceptibility to situations presents yet another practical problem for anyone wishing to consistently engage in moral behavior.

---

4 It should be noted that my decision to separate my discussion along these lines reflects an expository strategy, not necessarily a belief that motives and biases represent distinct kinds that can be cleanly separated in an accurate account of our psychology or neural circuitry. More obviously needs to be done in psychology and cognitive neuroscience before we will have the complete, correct account of how our motives interact with our cognitive biases and vice versa, but we need not wait until then to have a fruitful discussion.

2.2 Problematic Motives

2.2.1 The General Picture

If we wanted to make a claim about motives at the greatest level of generality, it would be fair to say that humans, like all creatures, are driven to survive and reproduce. What goes into surviving and reproducing on Earth, however, is a complicated matter. We have never been able to simply and directly choose survival over non-survival or having offspring over having no offspring. Our success in these endeavors has depended on our successfully fulfilling a complex web of interrelated subsidiary goals, goals that we must, then, also be motivated to fulfill.

Moving down a level of generality, we can say that we must be motivated, first off, to attain the resources out in the world that we require in order to survive and reproduce. The most basic and direct way that this works is very familiar. If I am thirsty, for instance, and there is water in front of me, I feel driven to drink it. If I am hungry and there is food available, I feel driven to eat it. If the elements are dangerous and uncomfortable, and there is shelter, I feel driven to use it. When these basic necessities are not simply available, then we must find some way of acquiring them, whether it means walking to the stream, finding a fruit tree, or something yet more complicated. We can sum this up by saying that we are generally motivated to discover and pursue successful ways of acquiring the resources necessary for our survival and reproduction.

One way of acquiring the resources that we need is by simply searching for and discovering resources out in the world. This benign acquisition is only possible in a situation in which resources are plentiful enough. When they are not, competition arises. We can
think of competition as existing on a spectrum from less aggressive forms of competition like being quick to gather food from a limited source, departing, and hiding, to more aggressive forms such as killing another creature and taking its resources, or opportunistically killing competitors.\(^6\) We pursue aggressive acquisition in all sorts of inter-species ways, like when we take honey from a beehive, collect a bird’s eggs, or kill an animal for food, and in intra-species ways, when we take things from other people. The general strategy of aggressive acquisition is a successful enough one that many animals, including we humans, have developed the aggressive motivations that push us to pursue it in a variety of forms, including forms that entail killing or maiming the owner of, or the competitor for, the sought after resource. The flip side of the general motivation to acquire resources at others’ expense is a general motivation to defend ourselves as well as the resources that we have acquired. The ability and drive to defend oneself and one’s resources has been in an evolutionary arms race with the ability and drive to take resources since long before there were humans. As resource predation has become more cunning and more aggressive, as has resource defense, and there is very good reason to believe that this has helped shape the general structure of our modern human motivations.

In addition to the aggress/defend dichotomy, a key strategy that creatures like us have adapted to get the resources they need is cooperation.\(^7\) Cooperating makes possible a whole range of difficult but rewarding resource-acquisition tasks that would be impossible


\(^7\) This is not, of course, to imply that the strategy of cooperation, and the development of pro-social motivations is somehow causally or otherwise independent of the pattern of aggressing/defending. In fact, cooperation seems to enable both better aggressing and better defending.
for a single individual to accomplish, such as the hunting of an animal too unwieldy for a
single individual to hunt successfully. It also often serves the same risk-pooling function
that is played by, say, homeowners insurance. Group foraging arrangements ensure that if
somebody has a few bad days in a row, he or she will not go hungry as a result. Because of
the survival and reproductive benefits of cooperation, many animals, including humans, have
adapted a variety of motivations that lead us to act in pro-social ways, driving us to live in
cooperative societies with one another.

When creatures cooperate to the degree that we do, they in turn create a new sort of
environment for themselves, one in which new strategies are successful or unsuccessful.
When different strategies are successful, different motivations that push creatures to enact
those strategies survive and proliferate. As such, the way in which our general motives to
reproduce, acquire resources, and defend ourselves are expressed in our psychological
makeup is through motives that are specific to our complex social environments. Many of
these motives drive us in a pro-social, virtuous direction. As I will discuss in the next
section, many do not.

---

8 A related complication to bear in mind is that the more specifically we describe a behavior, the less likely it is
that humans in general are motivated to perform it. While a general motivation to amass a cache of food may
be universal, the motivation to find any given specific food is not. Yet another complication that is perhaps
worth noting is that when we say that a person has a particular motivation, we can say this at a number of
different levels of description. We can make a statement about motivation at, for instance, a general strategic
level, at a behavioral level, at a psychological level, at a functional level, at a neural level, and at a physical level.
“X is motivated to do Y” means remarkably different things at each of these levels of description. One goal of
this section is to get the discussion down to within the psychological, functional, and neural levels, because this
is where the interesting data that we most need to grapple with lie.
2.2.2 Problematic Motives: A Few

2.2.2.1 Self-interested Desires

Perhaps the most familiar, longest recognized, most basic, and least socially-mediated of our motives that serve as impediments to virtue is simply our drive to acquire the resources that we need, and the selfish orientation of that drive. We want a lot of things. Some of them, we do want for other people. Most of them, we want for ourselves. While examples permeate our daily lives, we can also see this quite directly in the way people typically play economic games like the dictator game. In the dictator game, a player is given an amount of money, and is instructed to divide the money between himself and another player. A recent meta-study of the vast majority of dictator games that have been played in laboratories across the world found that, on average, a player chooses to offer the other player only 28.35% of what he or she was given. This amount varies given a variety of conditions, including age of the players, how anonymous the players are, etc., and this is, of course, a mean that distorts the fact that some players give nothing, some players give all, and some players give an even split, but the upshot is clear: humans are generally motivated to take a bigger slice of the pie.

We can further see this when we look at studies that show the pervasiveness of strategies like free riding in the context of public goods investment scenarios. In a study in

---

9 It’s fairly clear that we have some altruistic motivations that chimps do not have. When given an opportunity to pull a lever that gave food to themselves as well as another chimp, chimps did not prefer the “benefit to others” lever. Children, however, readily chose in a similar study the option of a sticker for themselves and another child over the option of a sticker merely for themselves. See J.B. Silk, S.F. Brosnan, J. Vonk, J. Henrich, D.J. Povinelli and A.S. Richardson, “Chimpanzees are Indifferent to the Welfare of Unrelated Group Members,” Nature 437 (2005): 1357-1359; C. Thompson, J. Barresi and C. Moore, "The Development of Future-Oriented Prudence and Altruism in Preschool Children," Cognitive Development 12 (1997): 199-212.

which people were asked to complete a contribution table saying what they would contribute as an investment in a public good given the average contribution of others, about a third of people chose to free ride and contribute nothing at all regardless of the average investment of others.11 Half chose a strategy of conditional cooperation, where increased investment from others led to increased investment from themselves. Even these cooperators, however, chose to invest less than the average at each increment; the motivation to cooperate was still tempered by a basic selfish urge.

This is not, of course, entirely bad news. Earlier economic assumptions about human self-interest and rationality would predict that people, purely self-interested and rational, would give nothing in both games. Most humans, unlike *Homo Economicus*, are clearly equipped with a variety of pro-social motives that keep us from making purely rational self-interested decisions. Our psyche, rather than being a univocal self-serving machine, is a swirl of self-interested desires, other-interested desires, moral impulses, top-down inhibitions of these various responses, and so on. While this is a better situation than we might have had to contend with, it can still often place us in difficult situations if we aspire to be consistently moral. If, in a given situation, fulfilling our self-interested desires would mean somehow harming others, being unfair to others, or the like, and if they are too strong to be counterbalanced by other-interested desires, moral impulses, or our executive control, they can lead us in a counter-moral direction.

2.2.2.2 Dominance

Pro-social motives allow social animals to overcome the Hobbesian trap that their selfish desires might otherwise lead them into by enabling effective cooperation. In doing so, as I mentioned at the end of section 2.2.1, they helped to create a cooperative environment in which a new set of strategies had the potential to help ensure that individuals get the resources that they need. As a social environment began to determine which genes got passed on, various motives differentially survived that drove strategies that allowed us to navigate these social environments.

One dynamic that emerged is the formation and maintenance of dominance hierarchies. Existing in a socialized dominance hierarchy allows those at the top the access to a greater share of resources and mating opportunities, but without the cost of the continual fighting that it would normally take to win such spoils. Members of the social group simply recognize that those at the top are entitled to a greater share, essentially ending a potential conflict before it starts. This also benefits those who are on the submissive end of the hierarchy, because they still get access to resources through social cooperation, but without the great costs that would likely come from losing fights. One’s role atop a dominance hierarchy is typically achieved through combat and other activities that serve as proxies for combat intended to intimidate and display to each party who would win were combat to actually occur.

In primates, a number of behaviors typify interactions between those in the dominant position and those in a submissive position in a dominance hierarchy. In a classic study on dominance in monkeys, Abraham Maslow noted that the dominant individual in a
A pair of monkeys will take most of a limited food supply, perform nearly all of the mounting, bullying, and fight initiation that occurs between the two, exhibit a greater freedom of movement, more exploration, and exhibit little to none of the cringing, fleeing, or passivity under aggression, sexual or otherwise, that is exhibited by the subordinate.\textsuperscript{12} The behaviors were not simply stable within individuals, as different pairings led some individuals to be dominant who had been subordinate in another pairing, and to exhibit the behaviors of the dominant individual. In a companion study, it was observed that over time, dominance relationships change, as those who can eventually dominate an individual previously above them do so, and then exhibit the behaviors that typify the dominant position.\textsuperscript{13} It is rather clear, then, that primates track their place in a dominance hierarchy and are motivated to perform a set of aggressive actions to assert their dominance when doing so is called for.

As a result of our pre-human evolutionary endowment as well as our own long history existing in dominance hierarchies, humans also come equipped with cognitive capacities and motives that steer us through our dominance hierarchies. This general motive gives rise to a variety of different behaviors. Some of them include elements of our social interactions that often times occur without our being consciously aware of them. We tend to size people up on the basis of all sorts of minute behaviors and characteristics, from their posture, how much space they take up in their environment\textsuperscript{14}, and how elevated a space they


take up in their environments,\textsuperscript{15} to how they react when we look at them with a certain type of gaze.\textsuperscript{16} In doing so, we are reading cues that tell us who is dominant, i.e. what the likely outcome would be if we were to challenge somebody for their greater share of resources, and as it turns out, we do so rather accurately.\textsuperscript{17}

Our motivation to succeed in our dominance hierarchies is rather clear when we look at the link between self-esteem and dominance. Maslow, summing up his interactions with low-dominance people, says that such people feel

“uncertainty, lack of confidence, general inferiority, shame or lack of pride, weakness, general admiration and respect for others rather than for themselves, a feeling of being, in a very metaphorical sense ‘below’ others, of being looked down upon, of wanting to be someone else rather than oneself, of being dominated by others, of lack of faith in oneself and one’s abilities. They very often feel shy, inhibited, timid, unworthy, self-conscious, and embarrassed. Generally, they cannot be said to be satisfied with themselves.”\textsuperscript{18}

Recently, links between social dominance and self-esteem have been corroborated, with the caveat that social dominance has a greater effect on some people’s self esteem than it does on others’.\textsuperscript{19} Social acceptance appears to play a greater role in self-esteem than being dominant, but this may be because social acceptance is a necessary condition for dominance

and, as I noted before, because social acceptance without dominance is still a viable way to secure what one needs.

In addition to reading dominance cues and identifying, via our self-esteem, with our position in the hierarchy, we also come equipped with motives to perform behaviors that, if successful, would establish our dominance in a potential confrontation. Often, the behavior that is motivated is a relatively benign proxy for combat. In a recent study, for instance, it was shown that people with more dominant traits tend to avert their gaze from a subliminally presented angry face later than people who are less dominant. In more dominant individuals, something drives a reflex to stand one’s ground and stare back when presented with anger rather than to simply avert one’s gaze and disengage, which is taken as a sign of submission.20

The behaviors that we are motivated to pursue are, in many cases, though, significantly less benign. Researchers on school bullying now tend to see it as a strategy for gaining social dominance among one’s peers.21 When asked why they engage in bullying, the most frequent responses in one study were “to feel powerful”, and “to look cool”.22 Other studies have shown that there is a strong association between the intensity of bullying behavior and ones degree of social dominance and one’s degree of resource control.23 While the same study showed that bullying was merely sufficient and not necessary for social dominance and resource attainment, it was clearly a viable strategy. As such, the purposeful

aggressing against and tormenting of weaker, more submissive members of one’s peer group avails itself in human communities as a strategy for securing the resources that we are motivated to pursue.

In addition to the bullying we see among adolescents, we can see our tendency towards dominance manifesting itself in a number of problematic ways. Gang violence and street crime are easily seen as being born from general struggles for dominance in a community.\textsuperscript{24} Intimate partner abuse is essentially the enactment of a small dominance hierarchy similar to what Maslow observed in his paired monkeys. Racial prejudice, anti-egalitarian attitudes, and other forms of discrimination can also been seen as effects of a motive toward dominance, correlating significantly with a group of personality traits psychologists have labeled the Social Dominance Orientation.\textsuperscript{25} Humans, some more than others, are motivated not simply to take what we want, and to engage in conflict when it becomes necessary to that end, but to seek out conflict and to aggress against others in order to establish their position. The general motivation to dominate thus adds an extra layer of motivations that steer people away from what we could consider morally correct behavior.

2.2.2.3 Punishment

When a creature exists in a society with other creatures who are motivated by self-interest, it is often a necessary strategic move for it to disincentivize the sort of self-

interested behavior that others are tempted to perform, lest the scheme of cooperation
crumble. Long before we (in principle) gave a monopoly on violence to the state, we had to
be the ones who doled out punishments to keep each other in line. When we add to that
fact the kind of reputation-tracking that a dominance hierarchy introduces, wherein one
individual freely aggressing against another is a sign to all about where they stand, we can see
an even greater need for early humans to disincentivize behavior that was decidedly against
oneself. Humans have thus adapted a variety of reactive attitudes, emotions that drive us to
respond to certain behaviors in certain ways, to seek out certain outcomes, typically
involving a harm of the perpetrator, when we feel that we or one of our group has been
wronged.

Our motive to punish is one of the many motives that keep us from being the
rational *Homo Economicus* that we might have been. In studies of ultimatum bargaining, in
which one player gets to choose how to split an amount of money and a second player can
either accept the offer or reject it so neither party gets anything, it has been shown that, even
in a one shot game in which doing so could have no effect of bettering the outcome for
themselves, people will reject offers that are too unfair.\(^{26}\) Furthermore, it has been shown in
ultimatum games and public goods games that when given the option of paying to punish
those who made unfair offers, attempted to free ride, etc., people have been willing to do so.
We want to punish those who are unfair, sometimes more than we want that extra dollar.\(^{27}\)
This makes sense, considering the effect punishment has in these games in inducing fairer

\(^{26}\) See Daniel Kahneman, Jack L. Knetsch and Richard H. Thaler, "Fairness and the Assumptions of

\(^{27}\) See Ibid., also Ernst Fehr and Simon Gachter, "Cooperation and Punishment in Public Goods
offers or greater cooperation. We are motivated to punish because it works to foster cooperation that ultimately makes us better off than we otherwise would be.

While this punishment motive does enable cooperation, and so spurs free riders and unfair bargainers toward moral compliance, it has also saddled us with the moral burden of being motivated, in a variety of situations, to harm others. The motive to reactively harm so as to punish is a blunter instrument than we may have wanted; it does not work according to what we would reflectively decide is a proper response to a harm. In being motivated to punish, we are not justice-seekers so much as reactive aggressors. A recent study on the neural activation patterns when a player chooses to exact punishment at a cost to herself was consistent with reactive aggression rather than experiencing some sort of inherent reward. We do not like punishing per se, but we are driven to do it in much the same way that we are driven to fight off an attacker.28

While choosing to pay to punish an unfair offer in a public goods game is fairly benign, our tendency toward reactive aggression qua punishment can easily be problematic in situations that are less neat than a public goods game. When we feel wronged, we often want to do something about it, and it is not difficult for us to feel wronged and want to seek revenge when we have not actually been wronged, to want to do something that is disproportional to the harm we have suffered, or to react before reflecting on the nature of our reaction. Instances of reactive aggression both mild and severe permeate our social experiences and “can encompass verbal and physical assaults, road rage, domestic and

---

workplace violence, and homicide”.\textsuperscript{29} Interestingly, the degree to which this motive toward punishment manifests itself in reactive aggression is, to a certain degree, culturally mediated\textsuperscript{30} and dependent upon the prevalence in a person’s system of hormones like testosterone and cortisol.\textsuperscript{31} What this tells us, of course, is that the punishment motive, like the dominance motive, is not a monotone human universal so much as a programmed strategy that certain people are wired to follow more strongly than others.

2.3 Problematic Cognitions

In addition to the motives that drive us, we have of course developed a variety of cognitive abilities that helped our ancestors in their attempts to survive and reproduce. It is by now very familiar that our growing ability to use language, to accumulate knowledge, to understand causal relationships, to perform logical reasoning, to override impulses, to choose courses of action and several other faculties were keys to our ancestors’ success. The imperfect process that cobbled together our current motivational system, however, is also responsible for our current cognitive systems. These systems have allowed us to do remarkable things, of course, but their limitations, too, can be problematic for those who wish to become virtuous.

One of our most distinctive and amazing features is our capacity for cognitive control, our ability to inhibit our responses and to choose behaviors that run counter to our initial inclinations. It gives us a behavioral flexibility that no other animal has, and offers, to some degree, the promise of a life in which we can choose moral behavior over whatever immoral behavior we are motivated to perform.

One downside to cognitive control, though, is that it is a quite limited resource, in a couple of ways. First it is easily taxed, easily depleted, and easily overwhelmed. Second, it is only a limited part of the way in which our cognitions unfold. Rather than reasoning through problems and deciding on particular courses of action, we tend to rely on various shortcuts and heuristics, and while these heuristics are often beneficial and clearly served to help our ancestors survive, they can lead us morally astray. I will discuss a few of the limitations of cognitive control and conscious reasoning first, and then move on to discuss some problematic heuristics that we use.

2.3.1 Cognitive Control and Ego Depletion

One of the most important aspects of cognitive control in helping us to ensure that our behavior consistently remains moral is our response inhibition faculty. This is our ability to assess and temper our emotions, urges, drives, etc. so that we can choose a particular course of action. We often have conflicting motivations, and these executive functions help us to weigh our various motivational inputs and determine for ourselves what behavioral outputs we deem correct, which in turn can help us to do what we ought to do rather than what our natural responses to situations would have us do. It has been shown, though, that
this response inhibition faculty, essentially what we call willpower or self-control, can be depleted.

Roy Baumeister's research program in what he calls “ego depletion” began with a study in 1998 in which, through a series of experiments, he attempted to show both that self-control was a unified faculty that acted in several domains and that the use of that faculty wore it out, like a muscle. 32 In each of these experiments, Baumeister et. al. put experiment subjects through a task designed to draw on their effortful self-control, and then put them in a subsequent situation that would also draw on their effortful self-control, and found that compared to control subjects, subjects who went through the initial effortful self-control task performed worse than control subjects. In the first experiment, subjects who were instructed to eat only radishes when presented with both radishes and chocolates later gave up far earlier than control subjects when asked to solve difficult puzzles. In the second experiment, the experimental subjects were asked to choose between presenting two different persuasive speeches while control subjects were merely assigned their speeches, and those in the choice condition showed a similar lack of persistence in puzzle solving. In the third experiment, engaging in an emotion suppression task led to poorer performance in solving anagrams compared to control subjects. In the fourth experiment, subjects depleted through a strenuous rule-following exercise took longer to quit watching a boring movie when doing so required an active choice than a passive choice, whereas those who had not been depleted took longer to quit when doing so was the result of taking a passive option.

Further studies have focused more directly on the implications of these experiments on the moral quality of our behavior. One study showed that those participants who had been put in a situation in which they had to expend self-control resources responded to subsequent insults from a purported partner in an experiment with greater aggression than those who had not gone through the self-control task. In one experiment, subjects were depleted through a similar food-temptation condition, and then received harshly negative feedback on an essay from their “partner”. They were then to prepare food for their partner, who they were told did not like spicy food, and the food was to include “adequate hot sauce”. Those who were depleted added more hot sauce than controls. A second experiment showed that ego-depleted subjects were more willing to blast experimental partners with white noise when given an opportunity to do so in a competitive game, and also introduced experimental features that determined that ego depletion was not itself what spurred the greater anger or aggressive response. In a third experiment, those who had gone through self-regulatory depletion and were then given harsh feedback gave far worse scores on a candidate evaluation form for their partner’s application for a research assistantship than those who had not been depleted.

In addition to raising the chances of aggressively responding to insults, self-regulatory depletion also makes people less likely to feel guilty and, in turn, to offer others help. It makes it more likely that people will be make choices that make dishonesty and

cheating more likely, as well as increasing the likelihood that they eventually will cheat. One study found that the overclaiming of correct answers by depleted subjects who were allowed to score their own tests, and who were going to receive a monetary reward for their performance, was 197% higher than it was by their non-depleted counterparts. People who had been depleted were found in another study to be more susceptible to persuasion than those who had not been. While this is not a direct instance of moral failure like greater dishonesty, decreased helping, or increased aggression, it is fairly easy to see how it could get some people in trouble in certain circumstances.

The limitations of this resource are not exhausted merely by noting that it can be depleted, of course, for even in an undepleted state, people are often unable to override various impulses, force themselves to attend to morally relevant aspects of their environments, etc. Some motives are simply too strong for some people even on their best day. It is one thing to be able to keep oneself from laughing during a funny movie. It is quite another to keep from lashing out at somebody who just spilled coffee on one's favorite shirt. Given the effects of depleting it, cognitive control clearly helps people to counteract their natural but counter-moral urges, but even when it is operating well, it is still no silver bullet.

---


2.3.2 Biases

It is bad enough that our active ego can be so easily depleted, leaving us at the mercy of our habits and inclinations. One thing that makes our situation even worse is that even when we are using our active ego, trying to assemble our perceptions of the world into considered judgments, reasoning about what to do, and choosing courses of action based on our reasoning, things can and do often go wrong. This is because a good deal of our thought about the world is informed, at some point along the information processing stream, by biases that are built into how we perceive the world, attend to facts and details, respond to evidence, respond to arguments, make judgments, remember facts, form opinions, and make decisions.

The accuracy of our perceptions and our reasoning are often morally important, because a good deal of what we need to be able to do in order to judge which behaviors we ought to perform includes correctly assessing the situations that are presented to us. If we perceive, assess, evaluate, and reason about the world, especially the social world, in ways that tip the balance away from accurate judgments of our situations, then even when we feel like we mean well, we may end up doing things that we would recognize to be morally incorrect responses to our actual situations.

Certain biases are, of course, more likely to present an impediment to moral behavior than others. This is especially true in light of the motivational structures that I discussed in section 2.2.2. If we are motivated to, say, reactively harm others as a form of punishment when some situation type S obtains, and we are biased toward interpreting our situations such that we are likely to think we are in situation type S more often than we actually are,
then our problematic motive toward reactive harm will see the light of day even more often. Likewise, if we are motivated to share our resources with those who are deserving, but we systematically misperceive people’s actions such that members of our ingroup are always seen as more deserving than members of an outgroup, then this may be another practical obstacle between us and the behavior that our morality dictates.

2.3.2.1 Biases: A Few

Human decision-making is a complicated process. For any given psychic economy in any given situation, the addition or subtraction of some cognitive bias might yield a better or worse moral result depending on the details of the situation and details of the person’s psychology. Furthermore, it is going to be a long and difficult process for us to get a complete description of every cognitive bias that affects people’s reasoning, especially because different people will be differentially affected by different biases. With those caveats in mind, I do think that it is possible for us to give an account of biases that are both widespread, robust aspects of human psychology and, with a significant probability, interfere with people’s attempts to live up to their moralities such that attempting to counteract them will have net-positive moral effects. Here I will discuss a few biases noted in the psychological literature and why I think that attempting to overcome them will have a beneficial impact on people’s moral practice. These biases I will discuss, of course, are not specific individual weighted cognitions; to list those would be an impossibly specific task. They are rather general terms for collections of individual weighted cognitions that share some characteristic, typically a shared effect on our higher order thought and behavioral
processes. For instance, the first bias that I will discuss, the self-serving bias, is a name given to “any cognitive or perceptual process that is distorted by the need to maintain and enhance self-esteem”,\(^{37}\) which can obviously occur in too many different ways to list. For this reason, I will simply refer to the collection of self-serving biases.

2.3.2.1.1 Self-Serving Biases

The self-serving biases steer our cognitions about our selves and the events in our world toward the end of a positive view of ourselves. The way in which they manifest themselves is mostly in a tendency to attribute causal efficacy to positive facts about ourselves whenever we experience positive outcomes and attribute causal efficacy to external sources whenever we experience negative outcomes. A recent meta-analysis of experiments designed to test whether or not such an attribution bias exists “yielded an overall weighted mean effect size of .96, a very large effect indicating that in general, people make more internal, stable, and global attributions for positive or success events than they do in negative or failure events”\(^{38}\).

The reasons we are affected by these biases are rather clear, given our history existing in dominance hierarchies, the relationship between self-esteem and perceived place in the hierarchy, the relationship between being higher in the hierarchy and fighting more aggressively for resources and mating opportunities. Dominance correlates with


reproductive success, self-esteem correlates with dominance, and the self-serving bias helps to keep the facts of the world from impinging on our self-esteem. Neuroscientific evidence has shown that the internal attributions we make when positive events occur happen more automatically than the external attributions we make, which suggests a simpler architecture and a more central role for those attributions. It has also been shown that when we make self-serving attributions, there is activation in the dACC and the dorsal striatum, which may “suggest that these decisions have a rewarding effect that is due to a reinforcement learning process.”

While these biases are globally robust, they vary in people according to their different locations, ages, and situations. They affect people of certain ages more so than people of other ages, with a peak in pre-adolescence that declines during adolescence and remains lowered until late adulthood. They are also more prevalent in certain societies than others, with Western societies as a whole displaying a greater effect overall than Asian societies as a whole. The same meta-analysis found that some societies displayed no self-serving attribution bias at all, including the Japanese and Pacific Islander samples. A different meta-analysis looked at studies across which 14 different factors were used to moderate the self-serving bias, and found that when the moderating factors led to a situation that was

---

40 Mezulis, et. al., “Positivity Bias”, p. 734.
41 Ibid. p. 735.
42 Ibid., p. 735. Something to note for later, this gives us hope that, although this bias is a sort of universal that we have to deal with, there may be cultural remedies to it.
43 W. Kieth Campbell and Constantine Sedikides, "Self-threat Magnifies Self Serving Bias: A New Meta-Analytic Integration," *Review of General Psychology* 3, no. 1 (1999): 23-43. Those moderators were (a) role (participants are actors or observers), (b) task importance (participants regard the task as important or unimportant), (c) self-esteem (participants have high or low global trait self-esteem), (d) achievement motivation.
more likely to threaten participants’ sense of self, the effect of the self-serving bias was magnified. This is consistent with the idea that the function of the bias is to preserve self-esteem, but also shows that its functioning is more situationally fluid than one might think.

The evidence of self-serving cognitions can be seen not only in attributions, but also in the way in which people construct narratives, remember events, and report their reasons. A recent study looked at the confabulations of amnesia patients, the false facts and reasons that they create to fill in various gaps in explaining their behavior to others, and found that the confabulations were largely pleasant and self-enhancing.\textsuperscript{44} Split-brain patients have also been known to claim to have reasons that they could not have to explain certain choices that they make as a result of stimulus that this hidden from their left verbal hemisphere.\textsuperscript{45} These kinds of confabulations are self-serving because they protect one’s self-image as an in-control and reasons-responsive agent. Just as in causal attributions about positive and negative outcomes, the goal of protecting our self-image as reasonable motivates the relevant reasoning and belief-formation processes. These self-serving confabulations and rationalizations also occur in non-clinical populations. Richard Nisbett and Timothy Wilson

\begin{itemize}
\item (participants have high or low achievement motivation),
\item (e) self-focused attention (participants are high or low in self-focused attention),
\item (f) task choice (participants choose or are assigned a task),
\item (g) outcome expectancies (participants expect task success or task failure),
\item (h) perceived task difficulty (participants perceive the task as easy or difficult),
\item (i) interpersonal orientation (participants adopt a competitive or noncompetitive–cooperative orientation),
\item (j) status (participants' status is either equal or unequal),
\item (k) affect (participants are in a positive or negative affective state),
\item (l) locus of control (participants' locus of control is internal or external),
\item (m) gender (participants are female or male), and
\item (n) task type (participants are involved in skills-oriented tasks or interpersonal influence tasks).
\end{itemize}

have shown that people often report making decisions for reasons that they could not possibly have and stick to their stories when asked. ⁴⁶

Whether it is the result of these self-serving attributions and our tendency to construct favorable narratives and reasons, or something that merely goes hand in hand with them, it has been shown in the psychological literature that people also tend to show a superiority bias, an overestimation of themselves with respect to the average in a number of domains⁴⁷, from abilities to character traits. In one study, “subjects perceived themselves to be characterized more by desirable traits than the average college student, and less by undesirable traits.”⁴⁸ A similar study found that “Nearly all of the individuals studied here displayed a pervasive tendency to cast the self in more positive and less negative terms than they portrayed most other people.”⁴⁹

Self-serving biases can have morally problematic effects that stem from how they influence our perception of our selves and what we deserve in comparison to others. The tendency to attribute our successes to ourselves and our failures to external sources and to overestimate our good qualities and underestimate our bad qualities in relation to others’ can, over time, lead us to believe that we are more worthy of assistance or aid than others, that our mistakes are more worthy of forgiveness than others, and that we are, in general, 


better people than those with whom we interact. This can have two important negative effects on our moral practice.

First, it can set up an inappropriate factual background for determining what to do in our interactions with others and in our participation in our communities. It gives us, essentially, an incorrect judgment of our true situations that may, in the manner I discussed in 2.3.2, lead us to perform actions that we would recognize as incorrect for our true situation. The nature of these incorrect judgments may also tip the scales in favor of morally inappropriate actions given our motive toward acting in reactively aggressive ways that serve to regulate others’ behavior toward us. If we are in a situation in which we are somehow failing, and our self-serving bias leads us to believe that there must be some external source, we are more likely to incorrectly think that (or feel that) we are being wronged by people and to then reactively aggress against them. If we are in a situation of conflict, and we are likely to unduly believe that we are in the right and the other person is in the wrong, similar morally incorrect behaviors may occur because the conflict will likely become exacerbated.

Second, it can obscure from us the need for further moral self-cultivation in a couple of ways. It can make us believe that we are more morally developed than we actually are, or at least, more morally developed than the average person, so we do not take our moral cultivation as seriously as we otherwise would. It can also lead us to reflect differently on situations in which we have failed to live up to our moral standards, perhaps situations in which we came into conflict with others like I mentioned above, as situations in which we were in the right and the other person was not, regardless of how we might respond to a neutral, third-person account of the same interaction or how we might describe the
interaction were we in the other person’s shoes. This can keep us from learning the appropriate lessons from our interactions with others, how we ought to treat them for certain outcomes to obtain, how to avoid morally hazardous conflict, etc.

It could be argued, though, that the self-serving biases, in protecting our self-esteem and self-concept, are adaptive, that they actually benefit people, and so attempting to remove them may simply be a road to low self-esteem and depression that will curtail or at least outweigh any possible benefits that removing these biases may have on our moral practice. It has been shown, for instance, that “the use of self protective strategies, such as the SSB, is associated with successful coping and mental health”\(^\text{50}\). It clearly does the job that it is intended to do in the short term, one that we might think would lead people to be more stable and thus more capable of interactions with others that live up to their moral standards. It also appears, though, that the self-serving bias, while it does typically correlate with lower rates of depression in general, is not necessary for these lower rates to occur. Asian populations show a significantly smaller bias effect, for instance, but the rates of depression in those populations are not significantly higher than what we find in the West.\(^\text{51}\)

The bias, then, seems an unnecessary path to the good that it does provide.

It has been shown, though, that in addition to being unnecessary to ward off depression, the self-serving bias does indeed have negative moral effects that we might


anticipate. Relying on self-enhancing biases as a strategy correlates in both men and women with exhibiting a range of anti-social behaviors, which makes sense given the impact we can see of downplaying one’s role in one’s social failures on one’s continual moral development. People also respond quite unfavorably to those who engage in self-enhancement after we witness them doing something poorly, something that can ultimately upset even the process of building a secure sense of self because it can lead to ostracism and greater self-threat. It seems rather clear that attempting to rid ourselves of the self-serving bias would be to attempt to get rid of a good deal of bathwater without risking any baby.

### 2.3.2.1.2 In-group Favoritism and Outgroup Negativity

The next family of biases that I would like to discuss are ingroup-outgroup biases, something we can see as relatives of the self-serving biases. In fact, the two most likely mutually reinforce one another, considering how clearly the protection of our images of our in-groups is related to the protection of our images of our selves. It has been shown, for instance, both that intergroup discrimination leads to increased self-esteem, and that threats to self-esteem lead to increased inter-group discrimination. As a result, we reason about

---

54 While a 1998 review found that there is not much support for the “social identity hypothesis” that the need for global trait self-esteem motivates intergroup discrimination, it did find support for the idea that successful competitive intergroup discrimination enhances social identity and thus increases the specific state self-esteem associated with that identity. See Mark Rubin and Miles Hewstone, "Social Identity Theory's Self-Esteem Hypothesis: A Review and Some Suggestions for Clarification," *Personality and Social Psychology Review* 2, no. 1 (1998): 40-62.
and perceive our groups in biased ways just as we reason about and perceive ourselves in biased ways. As one would expect, this includes a pro-in-group pattern of attribution and evaluation in which people attribute positive outcomes for their ingroup to stable dispositional characteristics of their group, and attribute negative outcomes to external sources;\textsuperscript{56} this ingroup-favoring attribution bias has been shown to exhibit as strong of an effect as the self-serving bias.\textsuperscript{57}

This ingroup-outgroup bias is remarkably easy to activate. It does not require, for instance, that the group membership be central to a person’s self concept in any way; in fact, the group membership can be temporary and arbitrary, so long as at least something is at stake for the group. It has been found that merely dividing people into groups had no effect on the members’ evaluations of members of their ingroup or their outgroup, but even flipping a coin to determine which of the two groups won a prize led to a significant pro-ingroup bias in evaluations of ingroup and outgroup members.\textsuperscript{58}

This bias in favor of the ingroup is found in such minimal groups not only in the attributions that people make, but also in people’s willingness to distribute resources in resource allocation games\textsuperscript{59} and to cooperate in games like the prisoner’s dilemma\textsuperscript{60}. This


occurs regardless of the variety of different strategic options that are available to players; although some games used in experiments may confound options like maximizing one’s own group’s total share and maximizing the relative difference between one’s ingroup and the outgroup61, one that attempted to unconfound these options found that the bias remained in general, although maximizing relative difference was less popular than either maximizing the joint take in favor of the ingroup or fairness, indicating that harming the other group was less motivating than benefitting one’s own group.62

The effect of our ingroup-outgroup biases is strengthened by our susceptibility to the accentuation effect, a cognitive tendency whereby, having placed things into categories, we exhibit a sort of confirmation bias, paying more attention to information that supports the categorization than we do to information that upsets the categorization, thereby enhancing the differences we see between categories by consistently tilting our evidence pool in favor of greater inter-categorical differences. One instance of this tendency, people have been found to remember people’s faces as more ethnically-stereotypical than they actually are.63

When we categorize people as members of an outgroup, then, we tend to see them in ways

62 See Bornstein, "On the Measurement of Social Orientations".
that accentuate their differences with us rather than their similarities. Given our tendency to see our ingroup as more positive, this accentuation of intergroup differences can lead us to pay more attention to the negative aspects of outgroup members. It has been found in experimental studies that people have a tendency to generalize more strongly from negative impressions of individuals from an outgroup to a characterization of the outgroup as a whole than they do from positive impressions of outgroup individuals.\(^{64}\) This greater salience for the negative traits of outgroup members is also shown in the way that people recall group members’ behavior, with people remembering significantly more negative behaviors of outgroup members than ingroup members.\(^{65}\) It has also been shown that people accentuate group difference in the direction of humanizing ingroup members and dehumanizing outgroup members, attributing fewer uniquely human “secondary” emotions to outgroup members than to ingroup members\(^{66}\) and attributing more non-uniquely human “primary” emotions to outgroup members\(^{67}\), indicating an important empathy gap between ingroup and outgroup members in addition to the evaluative gap.

With these sorts of effects, it is not difficult to see how the ingroup-outgroup biases can have a negative impact on a person’s moral practice. Just as in the case of self-serving biases, they can make us more likely to feel as though certain people are less worthy of good treatment, or more worthy of bad treatment, and this can obviously affect how we behave


\(^{67}\) Maria-Paola Paladino, et. al., "Differential Association of Uniquely and Non Uniquely Human Emotions with the Ingroup and the Outgroup," *Group Processes and Intergroup Relations* 5, no. 2 (2002): 105-117.
toward others and how we respond to the treatment or mistreatment of others. They can make us less capable of understanding and empathizing with the struggles of an outgroup’s members. They can lead us to distribute the resources we control in unfair ways and to support social policies that distribute resources in unfair ways, keeping members of historically oppressed groups from the opportunities that others enjoy. Because of these negative affects on moral practice, it appears that some way of dealing with the ingroup-outgroup biases may be something to desire in system of moral self-cultivation.

One complication here is that while ingroup favoritism and outgroup negativity are clearly related, they are not simply two sides of the same coin. It has been shown that outgroup aggression and ingroup favoritism are to some degree, independent. Something aside from mere ingroup favoritism, such as conflict for resources between groups, is needed for outgroup negativity to turn into outgroup aggression, so eliminating the pro-ingroup bias may not be a necessary component of eliminating the worst moral wrongs that are associated with group conflict. That said, there are many ways to fall short of one’s chosen moral standards as a result of ingroup biases without one’s behavior rising to the level of engaging in outgroup aggression.

A more troubling complication is that one of the functions of ingroup favoritism is that it leads to better ingroup cooperation. In one study, a game was used wherein players were given money and could contribute, at personal cost, either to a pool that benefited their ingroup, or to a pool that harmed the outgroup. People were vastly more likely to contribute to the former pool than the latter, showing that ingroup benefit is a much stronger motive.

---

than outgroup harm.\textsuperscript{69} Given that men tend to exhibit a stronger ingroup bias than women, another study tested whether the male coalitional psychology was founded on ingroup love or outgroup hate. In that study, “Male, but not female, participants…provided more money to the in-group than to the out-group despite the fact that they did not expect a more favorable treatment from in-group partners. At the same time, the in-group bias among men was found not to be a consequence of spiteful behavior toward the out-group.”\textsuperscript{70} This suggests that our tendency to form ingroups, rather than fomenting intergroup conflict, may be a resource in heading off intragroup conflict and creating the kind of interpersonal solidarity that leads to effective cooperation. The issue, then, from the perspective of creating or assessing a method of moral cultivation, is whether we can eliminate the negative effects of the ingroup-outgroup biases without losing the positive, pro-social effects of identifying oneself by way of group membership. This may be a situation in which there are tradeoffs for moral practice either way.

2.3.2.1.3 Racial Bias, Gender Bias, and the Halo Effect

There is support in cognitive science for the idea that our tendency toward self-serving biases and ingroup-serving biases is the result of our brain’s connectionist architecture; we make positive associations with our self, and negative associations with things that compete with the self.\textsuperscript{71} Our connectionist architecture leads us to think in


\textsuperscript{71} Tim Vanhoomissen and Frank Van Overwalle, "Me Or Not ME As Source of Ingroup Favoritism and
terms of prototype categories and associations, which helps to explain why people are susceptible to availability biases and why people are more likely to engage in outgroup stereotyping when put in a situation in which they are under cognitive load and thus more likely to rely on heuristics. This tendency to think in terms of prototype categories and connected associations makes us susceptible to further morally relevant biases beyond self-serving and ingroup-serving biases.

One such bias is a tendency that people have to automatically perceive people with some set of positive qualities as having other positive qualities, and some set of negative qualities as having other negative qualities, something that has been dubbed the “halo effect”. Altering one quality of a person, such as how warm or friendly he or she is, can have an effect on qualitative ratings of other aspects of a person, such as appearance and mannerisms, or accent, without people realizing or even suspecting that this effect is taking place. Our global evaluations, it seems, do not come from a sum total of our individual attribute evaluations, but rather exist in a sort of feedback loop with them; we make global evaluations based on a handful of individual attribute judgments, and then those global evaluations color our individual attribute evaluations. This happens in a number of ways. When a person exhibits some negative behaviors, this can lead one’s other negative behaviors appear to others to be worse than they would otherwise appear. How people

---

72
dress can affect perceptions of their intelligence and academic ability.\textsuperscript{75} The purported attractiveness of an author affects evaluation of their work;\textsuperscript{76} in fact, unattractive people have been shown to be perceived as “(1) Less intelligent, (2) less popular, (3) less happy, (4) less likely to continue their education and obtain prestigious jobs, (5) more dishonest and unkind, and (6) more aggressive and anti-social”.\textsuperscript{77}

This bias in perception can, of course, lead to a bias in treatment, and this is where the negative effects on people’s moral practice can result from the perceptual and evaluative halo biases, by, once again, giving us improper determinations of what different people deserve, when people have wronged us, when people have wronged others, etc. It is no surprise, then, that studies have shown, for instance, that being conventionally unattractive leads people to earn less money\textsuperscript{78} and receive harsher sentences in criminal trials.\textsuperscript{79}

Racial and gender biases work in similar ways to the halo biases and can have similar negative impacts on people’s moral practice. People categorize each other in certain ways, which includes race and gender categories. The categorizations lead to a variety of implicit


associations, these associations influence which facts become salient in people’s perceptions, and this in turn influences the evaluations that the people make, which, of course, affects their behavior.

Recently, tests have been developed to determine people’s implicit associations with respect to categories like race and gender. One of the things these widespread use of these tests have shown is an “automatic preference for White relative to Black. For both name ($d = 0.71$) and face ($d = 0.88$) tasks, participants showed an automatic preference for White over Black.” More specific implicit association results have been obtained as well, such as in one study in which “participants held strong associations between Black and Guilty, relative to White and Guilty, and these implicit associations predicted the way mock jurors evaluated ambiguous evidence.” Strong associations have also been found between male terms and science and career oriented terms, as well as between female terms and liberal arts and family oriented terms. There is no need to give an exhaustive account of all of the implicit associations that have been found, just so long as we understand that they color our perceptions and behaviors in ways that can be morally problematic.

Another tool for testing implicit racial bias is the shooter bias test. In this test, people are presented with scenes in which a person is either holding a gun or something more benign, and the player has to determine whether or not to shoot. It has been shown that, in cases like these, people tend with more frequency to shoot at African Americans in

---

82 Nosek, Banaji, and Greenwald, “Harvesting Implicit Group Attitudes”
general, and especially to shoot unarmed African Americans at a far greater frequency than unarmed Caucasians.\textsuperscript{83} As one would expect, priming people in ways that make racial stereotypes more salient can exacerbate this bias.\textsuperscript{84}

In a particularly vivid example of the way biases such as this can differentially effect our treatment of people, a recent study\textsuperscript{85} had people look at two identical legal memos in which a number of errors had been placed, including minor errors of spelling and grammar, facts, technical writing, and factual analysis. The names and educational background on the memos were exactly the same; the only difference at all was the stated race of the author. The average rating for the memo was 4.2/5 when it was marked “Caucasian” and 3.1/5 when it was marked “African American”. The average number of the errors that were found in the “African American” memo was greater than in the “Caucasian” sample, particularly in spelling and grammar (2.9/7 vs. 5.8/7), and there were 29 suggested edits for the “African American” memo compared to the 11 suggested for the “Caucasian” memo. Such biases, obviously, can lead us to treat people in morally problematic ways.

2.4 The Problem of Situationism

In addition to the mass of motivations and cognitive biases that can lead us in a morally problematic direction, there is another problem that we need to contend with. This

is that a good deal of our behavior is influenced by elements of our situations without our being aware their influence or even, sometimes, aware of the situational elements themselves. If our behavior is so easily influenced by our situations, then this presents an obstacle to building the kind of behavioral consistency that is a prerequisite for our behavior being consistently morally acceptable.

Much has been said about situationism, especially in the context of its relationship to the viability of virtue ethics, so there is not really much need here to give an exhaustive account of situationist findings. I only mean here to show some of the various ways in which our susceptibility to various effects from our situations can serve as an impediment to consistent moral behavior, so I need not present everything that one would need in order to make, say, a sufficient case against the existence of robust character traits that exhibit strong cross-situational consistency, like the kind given by John Doris or Gilbert Harman. I do not need to establish anything beyond the practical upshot that, in a variety of situations, people who think that they should perform B in some situation might be influenced not to perform B by some aspect of their situation, unless they find some way to counteract that influence.

Of course, it is true that some general upshot about the relative efficacy of people’s internal dispositions versus that of their situations may fall out of recognizing situations as an impediment. However, the debate about what specific form that upshot should take in discussions of character traits, from Doris’ and Harman’s strong thesis that only local,

situation specific traits exist to the position taken by critics of Doris and Harman, like Gopal Sreenivasan, who argues that “situationism does not really address the question of whether people have ‘character traits’, properly understood”, need not concern us here. This is in large part because this debate hinges so much on how we ought to understand the notion of character traits, and whether the facts presented in the literature warrant claims about character traits as they ought to be understood in the context of virtue ethics. The goal of my project, though, is ultimately to provide an answer to the question of how we can come to consistently accord with our moral norms, a goal we can accomplish without needing to defend any particular conception of what a character trait is or to relate such a conception to the psychological literature.

It should also be clear that we need not vindicate global character traits in order to think it possible to successfully develop a method of cultivating consistent compliance with our norms, because developing global character traits is not the only possible way to ensure that compliance. If, for instance, we are so affected by our situations that we can only be said to have local, situation-specific traits, and yet we have designed a form of life that consistently places us in the right sorts of situations, then we can still live in consistent accordance with our moral norms. The position that I am taking in saying that situations affect us in ways that can impede consistently moral behavior is simply that we need some way to overcome their potentially negative influence; what this most likely means is that the overall method that we should pursue should contain some combination of methods for increasing the strength of our moral dispositions and our conscious control so that we can

---

exhibit greater cross-situational consistency (irrespective of whether this leads us to be worthy of being said to have global traits) and methods for designing situations such that our dispositions need not exhibit as much cross-situational consistency in order to effectively keep us on the right side of the moral line (which, given the general consensus that situations do have effects on us, is something that we can all agree will make a moral cultivation method more effective regardless of just how possible it is to cultivate dispositions that we would call global traits).89

As one would expect, given the degree to which we are wired to keep track of where we fit in and how to behave around the people that surround us, one of the aspects of our situations that affects our behavior is the social situation in which we find ourselves. One of the more morally troubling ways in which we are susceptible to situations was discovered in a famous experiment by Stanley Milgram90, in which people were told by experimenters to administer what they thought were serious, even potentially lethal, shocks to other participants in a learning exercise. The degree of compliance was far above what people would predict both of themselves and of other people. These results were replicated in

---

89 We can note also that if pursuing this kind of mixed strategy can yield consistently moral behavior, and if consistently moral behavior is largely what we are interested in when we talk about virtue, then we may even want to eventually uncouple talk of virtue from talk of individual character traits. Our being virtuous may be dependent on things outside ourselves, or perhaps moral virtue is more properly thought of as a quality of interpersonal systems rather than individuals. If we do not privilege the individual as the locus of virtue, then the reality or unreality of global traits will cease to be such an important factor in discussions of virtue and whether or not we can cultivate it. As will become clear in Chapter 5, this greater focus on the relational nature of human behavior and the pursuit of a mixed strategy are a part of what helps to make the Confucian method particularly effective.

further studies, with a recent study finding only slightly lower rates of obedience than Milgram found in the early sixties.\footnote{Jerry M. Burger, "Replicating Milgram: Would People Still Obey Today?," \textit{American Psychologist} 64, no. 1 (2009): 1-11.}

Another famous experiment has shown that one’s social situation, especially the role that one inhabits in a particular power structure, can lead one to behave in ways that one would not otherwise behave. In Phillip Zimbardo’s Stanford prison experiment,\footnote{Craig Haney, Curtis Banks and Philip Zimbardo, "Interpersonal Dynamics in a Simulated Prison," \textit{International Journal of Criminology and Penology} 1 (1973): 69-97.} students were divided up into guards and prisoners, and were told to enact these roles. Over time, the behavior of many of the guards devolved into wanton cruelty, and the behavior of the prisoners ranged from depressed passivity, to complicity in the guards abuse, to rebellion.

Some people have criticized the validity of the experiment, on the grounds that it did not faithfully replicate a prison environment,\footnote{Ali Banuazizi and Siamak Movahedi, "Interpersonal Dynamics in a Simulated Prison: A Methodological Analysis," \textit{American Psychologist} 30, no. 2 (1975): 152-160.} on the grounds that the behavior can be explained by people playing roles rather than the effects of power,\footnote{Peter Gray, "Why Zimbardo’s Prison Experiment Isn’t in My Textbook: The Results of the Famous Stanford Prison Experiment Have a Trivial Explanation," \textit{Psychology Today}, October 9, 2013. http://www.psychologytoday.com/blog/freedom-learn/201310/why-zimbardo-s-prison-experiment-isn-t-in-my-textbook} and on the ground that the subjects may have self-selected for aggression and authoritarianism.\footnote{Thomas Carnahan and Sam MacFarland, "Revisiting the Stanford Prison Experiment: Could Participant Self-selection Have Led to the Cruelty?," \textit{Personality and Social Psychology Bulletin} 33, no. 5 (2007): 603-614.} While I believe these are valid criticisms, and that we should not see the experiment as establishing too sweeping a set of ramifications, I think that the experiment still effectively shows the more modest point that certain social situations can have negative impacts on people’s behavior from a moral standpoint.
The effect of our social situations goes beyond where we sit in relation to other people in social structures. The mere presence of other people can have a dampening effect on the likelihood of people exhibiting helping behavior, something that has been called the bystander effect. A recent meta-analysis of bystander effect studies has shown that the bystander effect is a robust aspect of our psychology, although the inhibition of helping behavior is reduced in situations that are clearly dangerous and situations in which additional bystanders can provide necessary physical support (situations in which, for instance, there is an attacker).

In addition to our social situations, we are affected in a variety of ways by other aspects of our environments. There is a link, for instance, between high temperatures and aggressive behavior that has been demonstrated in laboratory experiments as well as studies of data on violent crimes. Much like heat, loud ambient sounds can affect people’s behavior, making them both more aggressive, and less likely to exhibit helping behavior. Pleasant smells have a positive effect on the likelihood that somebody will help. People are more likely to perform dishonest and self-interested behaviors when in a dimly lit room

---


101 Pers Soc Psychol Bull May 1997 vol. 23 no. 5 498-503
or when wearing sunglasses. We judge others more harshly when in the presence of messy environments or bad odors. Even being exposed to various words has an effect on people’s behavior. Interaction with words related to hot temperature can make people more prone to express aggressive thoughts and to perceive others as hostile. Similarly, being exposed to words like “aggressively” and “intrude” over the course of a word game can make people more likely to interrupt when somebody is speaking.

In addition to the effects of these external situations, certain of our internal states can affect our morally relevant behaviors and judgments. One study of judicial decisions found that the likelihood of somebody receiving a favorable parole judgment was about 65% when the judge had just taken a meal break, falling to near zero over the course of the judge’s shift before the next meal break, after which it would rise again to about 65%. Considering the role of glucose in helping to ward off ego depletion, a lack of cognitive control going into the lunch break may help to explain this; of course, one of the situations that one can be in is a situation of depleted self-control. In addition to our meal schedule, our moods and having them affected in certain ways by elements and events in our environment can have an influence on our behavior. In one landmark study, the primary

---

determinant of whether or not somebody stopped to help whether or not people found a dime in a phone booth or whether they were given cookies.\textsuperscript{107} This general result, positive mood having a positive effect on helping behavior, has been replicated in several other studies.\textsuperscript{108}

So, it is clear not only that our situations can influence our behavior in a number of ways, but also that the influence of situations is particularly robust as well as relevant to our being consistently moral. Since our situations have such an influence on our behavior, the project of determining how we can come to consistently accord with our particular moral standards must involve some strategy for dealing with the effect of our situations, whether that means finding a way to counteract them or finding a way to engineer, shape, or control them.


Chapter 3: Moral Cultivation in the *Analects*, Part 1: Is There a Normative Theory in the *Analects*?

3.1 Introduction

I argued in Chapter One that a psychologically realistic understanding of how moral cultivation can take place plays an important role in both normative theorizing and naturalistic ethical discourse and practice more broadly, and I proceeded in Chapter Two to draw a detailed, empirically supported picture of the impediments that face people who embark on a path toward consistently behaving in accordance with their moral norms. Over the next two chapters, I will present and defend an interpretation of the *Analects of Confucius* from which I believe we can draw a method of moral cultivation that will allow us to make progress in overcoming those impediments.

My basic claim throughout these two chapters will be that the best reading of the *Analects* is one on which the text is centrally focused on issues of moral cultivation. There are two things that I mean by “best read” here. First, and most importantly, is an interpretive claim, i.e. that the best theory of the meaning of the text is one on which Confucius and his students are primarily engaged in discussing a method of cultivation. The second is a pragmatic claim, that the discussion that we find in the *Analects* on my reading has a greater contribution to make to current ethical philosophy and practice than a normative theory that we may be able to pull out of the text.
The argument for the interpretive claim has both a critical and a positive dimension. The critical dimension involves advancing a number of objections to various interpretations of the *Analects* that attempt to see it as containing and advancing some type of systematized normative theory, i.e. some attempt to find a standard that underlies the Confucians’ normative commitments. My objections to these various “normative theory” interpretations should help us to better understand what is happening in the ethical discussions in the text, since seeing how these attempts at systematization go wrong can give us a clearer picture of how the discussions in the text do and do not hang together. These objections can also lead us to draw some general, albeit cautious, conclusions about the prospect for eventually reading the *Analects* as giving us a systematic normative theory, conclusions that should push us to explore what contributions the *Analects* can make to modern philosophical ethics beyond giving us another candidate normative theory. This exploration should also be made easier by a greater understanding of the various distorting influences that various unsuccessful attempts at systematization can have and what we see in the text when we try to remove these influences. The positive side of the argument involves constructing my own interpretation of the *Analects*, and arguing that, especially in light of my critical arguments, my interpretation gives us the most plausible reading of the text. I present the critical arguments in this chapter, and the positive argument in Chapter Four. The critical arguments here will include, first, a general argument that we have good reason to think that the discussions of the *Analects* preceded and in fact prompted the first instances of what we could call normative theorizing in the Chinese philosophical dialectic. It will also include
arguments against specific interpretations of the *Analects* that each, in their own way, attempt to read it as presenting some kind of normative theory.

### 3.2 The Critical Arguments: That the *Analects* Does Not Present a Systematized Normative Theory

Before arguing that we should not view the *Analects* as presenting a normative theory, I need to make clear what I do and do not mean in making such a claim. To make this clear, I need to say both what I mean by “a normative theory”, and what I mean when I claim that something is or is not presented in a text.

I take a normative theory to be some attempt to systematize our normative judgments and give an explanation of why they are correct (in the broad sense of “correct” that transcends realist/anti-realist boundaries), either by bringing them into consistency with each other, some general normative standard, or both, or by giving some account of normative correctness on which some set of seemingly inconsistent judgments all deserve to be correct.\(^1\) This account of a normative theory should allow us to draw a crucial distinction between the claim that the *Analects* contains a normative theory and the claim that the *Analects* at least contains some positive moral views. I intend to assert the former claim and not the latter, as my overall interpretive claim that the *Analects* presents a method of moral cultivation of course implies that the subjects\(^2\) and authors held some positive moral views

---

1. This last clause is to allow for the possibility of various particularist theories.
2. I use “subjects” here as shorthand for “the people portrayed in the *Analects* as portrayed in the *Analects*” (and not for “subject matter”). I am assuming, I think plausibly, that the subjects of the text and whoever compiled it had roughly shared purposes in either doing and saying what the text describes or in creating the text so as to portray people doing and saying those things; the text is fairly clearly not, for instance, intended as satire.
and that they were concerned with cultivating an ability to act in consistent accordance with
them. Without moral norms, there is nothing to cultivate, much less anything to motivate
the development of an approach to cultivation. Such a claim would also be an interpretive
non-starter, considering the number of places in the Analects in which somebody clearly
expresses a moral opinion.

Now that we know what we are looking for when we look for a normative theory in
the text, what does it mean to say that we can find it in the text? I see several possibilities
that each set differently demanding success conditions. First, we could mean that we have
sufficient reason to say that the text explicitly presents and defends some normative theory.
Second, we could mean that we have sufficient reason to say that the subjects and authors
having held some normative theory is the best explanation of what we find in the text, even
if that theory is not made explicit. Third, we could simply mean that the text displays the
clear makings of some specific theory, i.e. that some significant portion of the text says
things that are some combination of consistent with, clearly implied by, and/or presupposed
by the theory without saying too many things that are inconsistent with it. We need not
decide here which of these standards we should favor, as I intend to claim that each
interpretive theory that claims to find a normative theory in the Analects fails to meet even
that third, weakest standard.

---

3 I may as well at least say, here, though, that I agree with Bryan Van Norden that one need not be able to
identify in a text a term that perfectly corresponds to something like “normative theory”, “duty”, or “virtue”, in
order to say that the text presents a virtue theory, for instance. See Bryan Van Norden, “Virtue Ethics and
Confucianism,” in Comparative Approaches to Chinese Philosophy, 99-121 (London: Ashgate Publishing, 2003),
particularly the section on what he calls the “lexical fallacy”.
The first move in my argument for this claim will be an attempt to show, in section 3.2.1, that the *Analects* precedes various discussions in the pre-Qin period in which we find criticisms of the earliest Confucians on the ground that they failed to produce a unified normative standard as well as various attempts from later Confucians and Mohists to present and defend what clearly look to be normative theories. Establishing the point that the *Analects* preceded explicit normative theorizing could make redundant any individual arguments that the *Analects* is not presenting a normative theory of type X, so why do I still move on to make individual arguments of the latter type in sections 3.2.2 - 3.2.4? Because, in dealing with interpretive theories of a text, especially a text like the *Analects*, the weight of total evidence matters. To have strong arguments at both the text-focused and period-focused levels is to have more total evidence, especially considering how consistent arguments at each level can reinforce one another.

### 3.2.1 Normative Theorizing in the pre-Qin Period

Philosophical discussions generally tend to work a certain way. Somebody poses a question, people attempt answers to that question, people respond to those attempts, sometimes by raising objections, sometimes by asking new types of questions, and then people attempt to respond to those objections or new questions, and so on. Some theories, like, say, emotivism, only make sense after certain general questions have been posed, like questions about how terms get their meaning, how they refer, etc. Going on the assumption, I think unassailably plausible, that the pre-Qin corpus contains an overarching discussion that works something like this, we can attempt to highlight when certain
questions got raised, what answers different participants in the dialectic gave, what objections to those answers got raised, etc.

This serves as the basis for what I shall from here on simply refer to as my *dialectic argument*. My dialectic argument says that we have good reason to describe the project of the *Analects* as something other than normative theorizing because the best account of the dialectic as a whole is one on which these normative questions were raised after, and in response to, the Confucian project gaining traction in the warring-states world. This account is superior for familiar reasons drawn broadly from the literature on what makes for a good interpretation. Without getting too deep into a fine-grained discussion of interpretation, which could easily fill a chapter, I will say that my account of the dialectic at least has the virtue of being as impartially philosophically charitable to the participants in the dialectic, where “philosophically charitable” here means not necessary maximization of the truth of their statements, but maximization of the reasonableness, insight, care, coherence, clarity, creativity, etc. of their philosophical moves.

The first piece of evidence that supports the dialectic argument comes from the fact that Mozi, a philosopher who came after and criticized the earliest Confucians, made a point

---

6 You can think of this as charity on the matter of how proficient of a philosopher everyone involved in the dialectic is. I basically mean to invoke whatever it is that we philosophers typically bring to discriminations between "good philosophy" and "bad philosophy".
not simply to argue in favor of a particular standard, but to argue that there ought to be a standard at all:

“Master Mozi said: Those in the world who take care of their matters cannot lack models and standards. There are none who lack models and standards and yet can bring their matters to completion. Of even those officers acting as generals and ministers, all have models. Of even the hundred artisans taking care of their matters, they also all have standards. The hundred artisans make squares with their carpenter’s squares and make circles with their compasses… Now the great ones ordering the world, and their deputies ordering the great states, and yet lacking a standard that measures – this is not as good as the distinguishing of the hundred artisans.”

This need for thinkers of the time to even develop standards is elaborated upon elsewhere in the Mozi, in a way that makes it even clearer that the Mohists take the earliest Confucians not to have engaged in something like constructing a normative theory. This passage occurs in the beginning of the Mohists’ argument in favor of taking Heaven-Nature as a grounding standard for their morality, and claims that a precondition of correctly applying and acting on moral terms like “benevolent” and “morally right” is an investigation into what it is that makes something morally right. In other words, the Mohists again argue for the very idea that we need normative standards:

“If the noblemen (junzi) of the world today desire to do what is benevolent (ren) and morally right (yi), then they must examine that from which moral rightness comes forth.”

---


8 Mozi, 42/27/1.
In reading such passages, we can either attribute to the early, *Analects*-era Confucians a project concerned with elaborating a normative standard, and thus attribute to the Mohists either ignorance of the *Analects*-era Confucian normative theory project or an unfair style of discourse, or we can attribute to those earliest Confucians a different project, and see the Mohists here as arguing for the need to address a new question in the dialectic, namely the question of what our normative theory should be. Only the latter option seems to me to be duly charitable to both the Mohists and the earliest Confucians.

We can say something similar about a passage in which Mozi criticizes the expensive funerals defended by the Confucians. His argument involves a thought experiment about a culture in which the first born son of every family is cut into pieces and eaten, and in which, when somebody’s father dies, they would carry the widowed mother away and abandon her:

“How can it be that this is the true way of benevolence and morality? This is what is called treating their practices as norms and their customs as morality.”

In attacking the practice, he does not appear to engage any purported Confucian standard. He engages rather with a mindset of unreflective acceptance of social practices without an investigation into proper moral standards. The example vividly illustrates that tradition and considered moral judgments can come apart, so more must go into determining the standard of morality than saying “We have always held these things to be right”.

---

9 Notice that to claim that we ought to hold elaborate funerals, as the *Analects*-era Confucians mostly do, is to express a positive normative judgment, but an unsystematized handful of such judgments does not a normative theory make.
10 *Mozi*, 39/25/75.
Mozi does, however, argue against some possible Confucian standards. After establishing that we need a standard, he goes on to assess the possibility of parents, cultural learning, or *junzi*\(^{11}\) as possible standards, arguing similarly in each case:

“When all take their parents as a standard-model, what is it like? Those in the world who are called parents are many, and yet the benevolent are few. If all take their parents as a standard-model, this standard is not benevolent. If the standard-model is not benevolent, we cannot call it the standard-model.”\(^{12}\)

This might, by the same dialectical coherence reasoning that I have been advancing, give us reason to see the early Confucians as advancing a normative theory, one to which Mozi needed to respond. However, the fact that Mozi chooses a grab-bag of potential Confucian standards, I think, shows that it was not at all clear to Mozi what the Confucian theory could have been. He refutes what he takes to be the Confucians’ best potential options at that point.\(^{13}\) This, coming on the heels of the clear Mohist argument that we need standards in the first place, gives us further reason to think that the *Analects*-era Confucians did not articulate such a standard.

A different source of evidence for my dialectic argument comes from the fact that we can see, in Mencius, a Confucian response to the normative challenge raised by the early Mohists and Yangists. He is not fond of such disputation,\(^{14}\) perhaps because it was not a part of the earliest Confucian tradition that he inherited from the *Analects*-era Confucians.

\(^{11}\) This term is used for “noblemen”, but took on a use in the *Analects* to denote a person of high moral achievement. It is sometimes translated as “gentleman” or “exemplary person”. I explain in Chapter 4 how I see it functioning in the text, and will sometimes leave it untranslated in the discussion in this chapter, as translation can often obscure textual discussions.

\(^{12}\) *Mozī*, 3/4/5-6.

\(^{13}\) Which, we can see, was part of the impetus for Mencius to develop, later in the dialectic, a Confucian normative theory grounded in human nature.

\(^{14}\) *Mencius*, 3B9.
Despite his distaste for disputation, Mencius’ environment, unlike that of the Analects-era Confucians, is one in which such disputation has become necessary as a result of the Mohist arguments that normative theories are needed and the Mohist and Yangist attempts to present and defend their respective normative theories. Mencius’s discussion of the goodness of the “four sprouts” and our uniform natural response to a child about to fall into a well,\(^\text{15}\) and his arguments with Gaozi and Gongduzi over the character of human nature (xing) and whether morality was internal or external,\(^\text{16}\) can be seen as characteristic examples of Mencius’ attempt to ground a systematized Confucian standard in human nature.

Mencius, also, like the Mohists\(^\text{17}\) and Yangists\(^\text{18}\), invokes the authority of tian (Heaven/nature) as a way of grounding his standard (as tian is ultimately responsible for giving us our xing). This was the normative standard-maker par excellence in Mencius’ time, though it was eventually abandoned by the later Mohists\(^\text{19}\) and Xunzi\(^\text{20}\) (and was, on some interpretations, criticized by Zhuangzi\(^\text{21}\)). Mencius’ use of that in-vogue standard, rather than a defense of some pre-existing Confucian theory and/or justification clearly drawn from the Analects should signal to us that the construction and defense of a normative theory was something Mencius brought to Confucianism in light of its post-Analects prevalence.\(^\text{22}\)

---

\(^{15}\) Mencius, 2A6.

\(^{16}\) Mencius 6A1-6.

\(^{17}\) See the Tianzhi chapters, although a paradigm statement of the position comes from Mozi, 42/26/44.


\(^{19}\) While the later Mohist canons mention other Mohist doctrines, like inclusive love and the identification of “morality” with ‘benefit’, appeals to tian’s will are conspicuously absent.

\(^{20}\) See the Tian Lan 天論 chapter of the Xunzi.

\(^{21}\) For one such interpretation, see Hansen, Daoist Theory, 277-280.

\(^{22}\) Notice also that Mencius does not try to defend the possible standards that Mozi criticizes, which should signal to us that Mozi really was trying to head off whatever the Confucian theory might say is the standard rather than responding directly to some claim from the Analects-era Confucians to have some theory based in a
Together, these points all draw us toward an interpretation of the dialectic in which the earliest Confucians asked the first real questions of the dialectic, which boiled down to “How can we master these norms that we have inherited, seeing how compliance with them is falling off?”, Mozi interjected that we need to not simply accept traditional norms, but find a standard, one of which he proposed and attempted to justify, which led to other competing answers to the question of normative standards (and eventually, on my interpretation, to the Daoist metaethical criticism of that project). This interpretation of the flow of the discussion makes better sense of Mozi’s criticisms and Mencius’ aims than one in which we attribute to the early Confucians the goal of presenting and defending a normative standard.

3.2.2 A Basic Standard in the *Analects*: On *Dao, Li, and Yi*

One approach to claiming that the *Analects* presents a normative theory is to claim that some term in the text is intended by the subjects and authors to denote the basis of the normative theory, i.e. some sort of ur-value, ur-principle, or grounding standard. Three terms in the *Analects*, “*dao*” (way), “*li*” (ritual), and “*yi*” (morality/rightness), have all been treated as this kind of theory-grounding standard in the literature. I will argue here that there are problems for interpretive theories that take one of these terms as denoting the basis of a normative theory in the *Analects* that should keep us from accepting such interpretations.

---

Attempts to see the Analects as presenting a normative theory based on dao are understandable. This is because of both the prominence of the term in Chinese philosophy generally, and its frequent use in Daoist and Neo-Confucian discussions to refer to metaphysically interesting, and thus possibly morally-relevant, daoš, e.g. what has been called “the cosmic dao” in the literature.\(^{24}\) I will try to show here, though, that the dao-based normative theory interpretations of the Analects do not have much textual support, in large part because “dao” is so often used in the Analects in ways that clearly do not fit with that later, narrower usage to refer to the cosmic dao.

One dao-based account is presented by Yu-li Liu, who argues that “the dao has an objective character”\(^{25}\), and is what “determines human excellence”\(^{26}\). The dao, on her account, is related to de (excellence/power) such that de is what one has when one “has gained an understanding of the dao”.\(^{27}\) She argues that since “dao is the basis of de, and dao has its source in Heaven, it is an objective existence and it is a predetermined way for man to follow.”\(^{28}\)

The first premise of this argument, that “dao” refers to the standard that we follow via having de, is suspect because it relies on an understanding that the modern Chinese term for ethics is “daode”, and that this reflects early developments in Chinese philosophy that connected the terms. While this etymological story is true in the general sense, since the

\(^{26}\) Ibid, p. 75.
\(^{27}\) Ibid, p. 80.
\(^{28}\) Ibid., p.80.
Daoists and Xunzi clearly started using them in a connected manner,\textsuperscript{29} it is not true in the more specific sense that this connection was there in Confucianism from the start, as Liu appears to imply. \textit{Dao} and \textit{de} are never compounded in the \textit{Analects}, and so it is hard to believe that their eventual connection reflects a specific relationship between the terms or their referents that was a part of the content of the \textit{Analects} all along.

Liu's support for the second premise, that \textit{dao} is based in Heaven (\textit{tian}), relies heavily on referencing later Neo-Confucian theories. After presenting a passage in the \textit{Analects} that claims that Confucius did not speak about the Way of Heaven, potentially contradicting her reading, she looks to Fung Yu-lan for support, who claims that reading Chinese philosophy as a “this-world” philosophy gives “only a surface view of the matter”\textsuperscript{30}. To elaborate on what the deeper view would be, though, she relies on one scholar's description of Neo-Confucianism. She also cites for further support the Neo-Confucian Cheng Yi's view that \textit{tian} simply is the \textit{dao}. I do not agree with this argumentative strategy, however, as we have little reason to think that the Neo-Confucians were simply recounting the views of the earliest Confucians, or that philosophical usages or connotations of a term like “\textit{dao}” did not change significantly over a millennium and a half. Absent a good deal of specific support for seeing continuities, it is a leap to say that the \textit{Analects}-era Confucians must have thought X because the Neo-Confucians thought X.

Where Liu does use the \textit{Analects} for support, her argument relies on similar leaps:


\textsuperscript{30} Feng Yu-lan, quoted in Ibid., p. 67.
“The importance Confucius attached to the Way can be seen from his remark, ‘He has not lived in vain who dies the day he is told about the Way.’ (4:8) Used in this sense, the Way seems to cover the sum total of truths about the universe and man.”

It is not clear how this usage of *dao* points to her interpretation, unless a necessary condition for not having lived one’s life in vain is hearing about the sum total of truths about the universe and man, but it is not clear why that would be.

Looking to the text of the *Analects*, we see not only the passage Liu tried to explain away, i.e. the one denying that Confucius discussed Heaven’s *dao*, but also discussions using “*dao*” that undermine *dao*’s purported nature as a grounding standard for a normative theory. For instance, Confucius says in one passage to follow one’s father’s *dao* for three years.

Unless everyone’s fathers follow the same *dao*, which seems unlikely, this means he is advocating different *daos* for different people rather than simply advocating The (Heavenly) *Dao*. A student complains of not having enough strength to walk Confucius’ *dao*. This either means that his students took Confucius’ *dao* to be Heaven’s *Dao*, which would contradict the statement that Confucius did not talk about Heaven’s *Dao*, or it means that “*dao*” often referred simply to ways in general in the *Analects*, and not always to some specific cosmic Way. We are told that people who follow different *daos* cannot make plans together and that people can study together without walking the same *dao* and walk the same *dao*.

31 Ibid., p. 73.
32 *Analects*, 5.13.
33 *Analects* 4.20, 1.11.
34 Even if we assume that the text is assuming that everyone’s fathers follow the same *dao*, and the intention here was to advocate following the one right *dao*, then why choose such a roundabout way of saying to follow the one right *dao*? Also, if everyone’s fathers already follow the one right *dao*, what need would there be for normative theorizing?
35 *Analects* 6.12.
36 *Analects* 15.40.
without taking their stand together,\textsuperscript{37} which shows that the Confucians of the \textit{Analects} see a multitude of possible paths, not one objectively laid out, universal, inevitable path like Liu’s \textit{dao}. If “\textit{dao}” played the role in the \textit{Analects} that Liu says it does, then we would expect passages like these to tell us something more like “somebody on the \textit{dao} cannot make plans with somebody off the \textit{dao}” or “we should be taking the same \textit{dao}”.

We are also told that it is the person who can broaden a \textit{dao}, not the other way around.\textsuperscript{38} I think that this helps to show that humans can contribute to the character of a \textit{dao}, and that this serves to undermine a general reading of \textit{dao} as the sort of objective, predetermined path that Liu claims. To whatever degree such an objective reading of “\textit{dao}” is necessary for viewing it as a standard that can underlie an ethical theory in the \textit{Analects}, undermining that reading helps to undermine the idea that “\textit{dao}” plays the role of a standard.\textsuperscript{39}

Another notable attempt at interpreting the \textit{Analects} as presenting a normative theory comes from Herbert Fingarette in \textit{Confucius: The Secular as Sacred}. The account Fingarette gives could be considered either a \textit{dao}-based account or a \textit{li}-based account. According to Fingarette, what the \textit{Analects} give us is essentially a normative theory that amounts to following the Zhou \textit{li} (the prescribed social rituals of the Zhou dynasty). The \textit{li} comprise, to Fingarette, “the moral order”\textsuperscript{40} in the Confucianism of the \textit{Analects}, which he sees as a Confucian answer to the quest for the “ideal way”\textsuperscript{41}. The \textit{li}, then, on Fingarette’s reading,

\begin{itemize}
  \item \textsuperscript{37} \textit{Analects} 9.30.
  \item \textsuperscript{38} \textit{Analects} 15.29.
  \item \textsuperscript{39} Cf. Luo, “Ren-based Interpretation”, p. 127.
  \item \textsuperscript{40} Herbert Fingarette, \textit{Confucius: The Secular as Sacred} (New York: Harper and Row, 1972), p.35.
  \item \textsuperscript{41} Ibid, p. 17.
\end{itemize}
clearly serve as what I would call a systematized normative standard. We typically expect a normative theory to give standards that are meant to be universal, to tell us how we ought to be or behave, and to be grounded in some way. On Fingarette’s reading, the *li* fit that bill. Not only is following them “the only moral and social necessity”\(^{42}\), but this necessity is grounded in “the cosmic *dao*”.\(^{43}\)

I have already argued that while the “cosmic *dao*” notion was prominent in Daoism and Neo-Confucianism, this alone gives us very little reason to read it into the *Analects*. One might push back and say that we have some reason to see grounding in cosmic notions at work in the text, given that the term “*tian*” (Heaven) is a significantly cosmic notion, that *tian* clearly has a *dao* given the use of the phrase *tiandao* (Heaven’s Way), and that *tian* is mentioned in contexts in which it appears normatively significant. One such context is the claim that Confucius realized *tian*’s propensities at age fifty\(^{44}\) (a significant milestone in the autobiography of his cultivation). It seems to me that even if this passage suggests that understanding *tian* is somehow an important part of one’s cultivation, or even that the notion of *tian* is somehow relevant to the normative content in the *Analects*, “*tian*” is not linked with “*dao*” enough to justify reading “*dao*” throughout the text as implicitly short for “*tian*’s *dao*”.

One passage does clearly link them, though, with a border official meeting with Confucius and then telling Confucius’ disciples that “all under *tian* have long since lost their

---

\(^{42}\) Ibid, p. 57.
\(^{43}\) Ibid, p. 57.
\(^{44}\) *Analects* 2.4. I am grateful to David Wong for pointing out the need to reconcile such passages with my interpretive view here.
dao, and tian is going to use your Master as a wooden bell clapper”\textsuperscript{45}, implying that the dao Confucius is attempting to implement is tian’s dao, or at least a tian-ordained dao. If we take the border official’s implication as an accurate claim, then it would seemingly contradict the passage that clearly states that Confucius did not talk about tian’s dao.

In resolving this contradiction, which we should do if we can in order to make the most coherent sense of the text, I think that we have more reason to see the border official as expressing something other than what Confucius taught his students than we have to see the statement that Confucius did not talk about tian’s dao as reflecting something other than what Confucius taught his students, simply because the border official is in a position less conducive to being correct about such matters than the authors of the \textit{Analects}. We should thus resolve the tension in favor of accepting that Confucius did not generally talk about tian’s dao, even if the border official did. For instance, just as a person could tell an atheist that she is doing God’s work, the border official might simply be offering his own take on Confucius’ capacity for improving the world. Or, given Confucius’ strategically varying his statements for his audience, perhaps Confucius gave a metaphysical spin to his mission in his discussion with the border official that did not reflect the way that he talked with his students or the way he conceived of his own project. Either seems to me more plausible than an interpretation that takes the border official to be accurately representing Confucius’ mission as bringing about tian’s dao and takes the authors of the passage denying that Confucius spoke of tian’s dao to be somehow mistaken.

\textsuperscript{45} \textit{Analects} 3.24.
This means that, while there may be cosmic notions in the text like *tian*, and understanding *tian* may be an important part of Confucius’ developmental path, we do not have a solid reason to see Confucius as advocating a *tian*-ordained *dao*, and this leaves us with little reason to see “*dao***” in the *Analects* as a generally cosmic notion. This already gives us some reason to doubt Fingarette’s account. If the *Analects*’ normative theory amounts to the view that the morally right thing to do is to follow the *li*, and what is supposed to ensure that following the *li* is the morally right thing to do according to the *Analects*’ normative theory is the idea that the *li* are somehow rooted in the cosmic *dao*, but this grounding in a cosmic *dao* is out of place in the *Analects*, then it is not clear why we should take the *li* to have the kind of objective moral character in the text that Fingarette claims. Further doubt can be raised by looking at passages in which Confucius claims that the *li* have changed throughout the dynasties leading up to the Zhou rituals he advocates and will continue to change, agrees with a particular alteration to a ritual procedure, mentions that rituals and music developed out of practices of the common people, and warns against a social danger that results from being too scrupulous in observing *li*. While the *li* are clearly treated as important, and much of the *Analects* focuses on the good they can do for us, they are simply not treated like the overriding moral standards as Fingarette claims, nor as we would expect them to be if we were to try to claim that they comprise the normative theory of the *Analects*.

This brings us to “*yì***”, a term that has been most often rendered as “morality”, “rightness”, and “appropriateness”. Perhaps this term refers to the central standard of the

---

46 *Analects* 2.23.
47 *Analects* 9.3.
48 *Analects* 11.1.
49 *Analects* 3.18.
Analects’ normative theory. D.C. Lau has argued that “in the last resort yi is the standard by which all acts must be judged while there is no further standard by which yi itself can be judged.”50 Lau’s argument for treating yi as an overriding normative standard comes in part from arguing that ren is the highest virtue in Confucianism, but “does not carry its own moral guarantee”.51 If ren does not carry its own moral guarantee, then something must, and what does is, to Lau, yi.

The first thing to notice about the line of argument that Lau advances is that in order to secure its conclusion, it must assume at the outset that Confucius has at least some normative theory worked out, and has further worked out that one of the moral terms we find in the Analects’ must be basic. This assumption forms the background conditions under which it is true that if his standard was not something other than yi, then it must be yi. This argument may work in a debate about what the theory’s standard must be if there is one, but it is question-begging in a debate against someone like myself who thinks that something else was going on in the Analects aside from articulating a normative theory.

When we look to the text to see how “yi” is used, we notice that it does work as a term for morality, or rightness, or appropriateness. This, however, does not mean that it denotes any particular standard or otherwise signals the existence of a normative theory. After all, people can be competent users of English and say all sorts of things using the terms “morality” or “rightness” without ever having attempted to systematize their moral judgments or to develop a conception of their basic moral standards.52 If yi really does play

51 Ibid.
52 The “moral dumbfounding” studies by Jonathan Haidt are clear indications that people can make moral
the role that Lau says, we should expect to see not only frequent use of “yi”, discussion of its importance, etc., but also some indication that it denotes some last resort, basic, standard that serves to systematize the various moral judgments we find in the *Analects.*

So is this what we find? Well, on a positive note, there are several places within the *Analects* in which *yi* is presented as an acceptable guiding priority for people to take under consideration in their dealings in the world. For example, in explaining the difference between being known (*wen*) and being great/prominent (*da*), Confucius says, “Those who are prominent…seek after what is most appropriate (*yi*).” We see the recommendation to be *yi* in yet another passage, when Zizhang asks about accumulating excellence, and Confucius replies that “To take doing one’s utmost (*zhong*), making good on one’s word (*xin*), and seeking out what is *yi* as one’s main concerns, is to accumulate excellence.” Even this passage, though, in promoting *yi* as an aim, simultaneously gives us a couple of reasons to be suspicious of the interpretation in which *yi* serves as a basic standard of some possible normative theory in the *Analects.* Not only is *yi* presented here as a mere part of some broader project called accumulating excellence, but it is also presented on equal footing with *zhong* and *xin*, themselves evaluative or virtue terms that should be subsumed under the normative theory’s basic standard if indeed one is being explicated.

In some discussions that mention *yi*, it is posited as the acceptable aim of the *junzi*, but while *yi* is given fairly high priority in these discussions, the discussions still lack an

---


53 And this would not be difficult to do in the language of the time, either. Mozi clearly lays out the need for a standard, argues for its role in resolving disputes, and puts it to use.

54 *Analects* 12.20.

55 *Analects* 12.10.
explanation of what yi amounts to. For example, when Zilu characteristically asks whether boldness or yi should be given first priority, Confucius responds that a junzi gives yi first priority. Confucius then explains the perils of being bold without paying attention to what is right or appropriate in a given situation. While Lau takes this as evidence that yi is the highest standard, I would argue that we can only draw the more modest inferences that to be yi is not identical to being bold, and that being yi is a more important concern. In another passage, Confucius states that “Junzi understand what is yi; petty persons understand what is of personal advantage.” Considering the way in which the junzi is upheld as a morally-successful person, this passage easily lends support to an interpretation of “yi” as the Confucian or, more generally, Chinese equivalent of “the right” or “the appropriate”. This passage clearly urges people to hold yi as a principal aim in their conduct, rather than holding personal advantage as such an aim.

But, while “yi” does seem to denote some concept of morality or rightness that should be a concern and focus for a Confucian, what we find is also consistent with reading the Analects as making exhortations to be moral and to be concerned with being moral that can be made absent (or at least prior to) any significant progress in an attempt to build a systematic conception of what it means to be moral. We can think of the subjects and authors of the Analects as having and using a morality concept, but one for which they have

---

56 Analects 17.23.
57 Going further than the points I make here, Shirong Luo argues that yi in this passage is treated as a virtue rather than a standard of conduct, that we are not justified in translating shang here as “supreme” (or “first priority”) and that yi is simply higher than boldness (yong). I disagree with his “virtue” reading of his “virtue” reading of yi, but agree that “junzi yi yi wei shang” is at least ambiguous between “junzi take yi to be highest” and “junzi take yi to be higher”. Of course, if he’s right on this either point, it helps my overall case. See Luo, “Ren-based Interpretation”, p. 135.
58 Analects 4.16.
59 This type of general exhortation toward morality also occurs in Analects 14.12, 16.10 and 19.2.
not yet laid out a conception that functions discernibly like a normative theory. Yi may be a
top priority for the Analects-era Confucians even if they have not significantly systematized
their normative view, just as morality today is nominally the top priority even for people who
have not significantly systematized their moral judgments.

Perhaps, though, since “yi” appears quite likely to be what we can call a morality
concept, we could better understand “yi” by going through the text and piecing together all
of the normative commitments expressed in the text. We could then discuss how
susceptible these normative commitments are to systematization, and systematize them as
well as we can in order to have an elegant way of explaining what the Confucians of the
Analects thought of as correct behavior; this might give us insight into how we should
understand the term “yi”, by telling us what the Confucians of the Analects thought of as
behavior that deserved to be called “yi”. This would certainly be a worthwhile project for
the intellectual history of Chinese thought and possibly for ethics, but I think that the results
of that project would be irrelevant to the position I am taking here, i.e. that “yi” does not
play the role of a foundational standard in a normative theory in the Analects. I think the
results are irrelevant because I think that our ability to thusly triangulate the content of “yi”
in the Analects does not speak as strongly against my position as the dearth of discussions in
the text clearly laying out its foundational role as a standard or the content that would allow
it to play that role speaks in favor of my position. This is in part because even if we are
capable of producing a quite systematized reconstruction, it strikes me as more plausible to
think that our ability to do so simply reflects a decent amount of consistency in their views at
the time than it is to think that it reflects a process of systematization that was important to
the project of the Analects-era Confucians but that they decided not to discuss in the text.

There is, however, yet another possibility to reckon with. Perhaps the text would resist systematization because what “yi” denotes is not a stable standard so much as a highly
contextualist or even particularist conception of rightness. Reading “yi” as a highly
contextualist standard would also go a long way toward showing that the Analects contains a
systematized normative view, even if that view is anathema to explication via principles.

There is some evidence in favor of this view. In Analects 4.10, Confucius says that junzi “are
neither bent on nor against anything; rather, they go with what is yi”. Because going with
what is yi is, in this passage, contrasted with having commitments to do or not do certain
things, this passage appears to imply that what is yi is not something that we can fix
abstractly, which, in turn, may imply that the underlying conception of morality at play in the
Analects is highly contextualist or particularist. This would also offer an elegant way to
explain why so many of the discussions involving “yi” read like mere exhortations to be
moral; simply telling somebody to do what is moral seems not to tell them what standard to
follow when what we expect is a clear set of rules, but it might in fact be as complete a moral
command as one can make if in fact “yi” denotes a kind of particularist or otherwise highly

60 We can think of the degree of context-sensitivity in a moral view as inverse to the amount of cross-
situational consistency there is in what are taken by the view to be correct moral judgments, e.g. how many
individual kinds of situational difference can affect, for instance, whether or not an act of lying is wrong. Views
deserve to be called highly contextualist when they exhibit an extremely low degree of cross-situational
consistency and they deserve to be called particularist when the judgments are thought to exhibit almost no
cross-situational consistency. Note that most moral views are fairly context-sensitive, though some, like Kant’s,
are less so.
61 Analects 4.10.
contextualist standard. Even then, I think that we have some good reasons to resist such a highly particularist reading of “yi” on the basis of what we find in 4.10.

While a reading of the passage as indicating that what is “yi” is influenced by context strikes me as plausible, there is another similarly plausible reading of the passage on which what one is not supposed to be bent on nor against is not any and every possible course of action, but rather the courses of action and/or policy proposals that his students, as aspiring officials, would have available to them in the course of their duties. On this reading, the passage advises something like avoiding getting attached to particular plans, but rather to always do what is right. This would be sound advice for students wishing to become morally upstanding officials even if what is right is not explicitly highly contextualist or particularist, because it serves to tell people to always take doing what is right as one’s concern. The idea that making this sort of exhortation, rather than explicating “yi” as highly contextualist or particularist, is the concern of the passage is supported by the fact that the subject of the passage is the junzi and how they are in the empire rather than yi and how it ought to be understood.

Even then, one could say that the passage does still clearly imply that what is “yi” must be context-sensitive to some degree. I agree that it implies this, but we should notice that most collections of moral judgments are finely-tuned enough that different contexts yield different results for similar types of acts. For instance, compare most people’s judgments of instances of lying and Kant’s judgments; the former display far more context-sensitivity than the latter. If even some of our sets of judgments display even low degree of

---

62 As I will argue in Chapter 4, this kind of strategy of keeping oneself ever-conscious of one’s goal to be moral is a part of the Confucian moral cultivation strategy.
context-sensitivity, keeping people ever-aware of doing what is right and keeping them from treating a course of action as always completely right or always completely wrong could serve as useful advice for people who want to make sure to do the right thing. Noting this allows the reading of the passage to reflect the clear implications of context-sensitivity of the passage, but it also allows it to fit neatly with my reading of other passages as general exhortations to make yi a priority. For these reasons, and also because the Analects contains some fairly clear, cross-situationally consistent moral judgments, like the judgment that sons should cover for fathers when they steal,\(^{63}\) there is not much support for seeing “yi” as denoting a highly contextualist or particularist standard. I believe that foreclosing this possibility, combined with the rest of what I have shown, sufficiently supports my claim that “yi” should not be read as denoting the foundation or content of a systematized normative theory in the Analects.

3.2.3 A Virtue Ethical Theory in the Analects?

Having argued that the Analects should not be read as advancing dao, li, or yi as a basic standard in some normative theory, I can discuss the possibility that the Analects contains a virtue ethical theory. Recall that, as I am only holding these interpretive possibilities to the weakest standard that I discussed in section 3.2, a successful interpretation of the Analects as containing a virtue ethical theory need not show that the subjects and authors of the Analects are either explicitly or implicitly committed to a virtue ethical theory. One must, however, establish more than that the subjects and authors

\(^{63}\) Analects 13.18.
discuss virtues or have a focus on cultivation or both, as discussions of virtue and cultivation can just as easily take place among people who say things that do not cohere with a virtue ethical theory, such as people who have no discernable unification in their moral view or people who explicitly think virtue terms are parasitic on other more fundamental moral standards. So, to clear that lowest bar, an argument that the Analects contains a virtue ethical theory needs to be able to show either that there is some character ideal, like “the virtuous person” that plays the role of a basic normative standard in the text, that there is some virtue or set of virtues that appear to constitute the ethical view of the text, or that some set of virtues at least serve as the basis for the total ethical view in the text, so that other standards, judgments, etc. are discussed in the text as deriving their moral status from the virtues. My contention is that there is no plausible way of showing any of these.

Some evidence in favor of the “character ideal” possibility is certainly available, in light of the use of the terms “sheng” (sage) and “junzi” (exemplary person) in the text. We might think that the sage works as a theory-grounding character ideal in the Analects, because the text does mention being a sage as a highly admirable state. When Zigong asks Confucius if the person who is broadly generous with the people and is able to help the multitude has

\[64\] Cf. Kim-Chong Chong, "Virtue and Rightness: A Comparative Account," in Conceptions of Virtue: East and West, ed. Kim-Chong Chong and Yuli Liu, 59-77 (Singapore: Marshall Cavendish, 2006), p. 60 where he remarks that “it might be held that from the ordinary person’s standpoint, sometimes, principles are important, while at others, the manner in which a person acts is paramount in deciding whether he or she has done the right thing or not”. I take “ordinary person” here to mean somebody with a pre-theoretical set of moral commitments. Cf. section 3 of this chapter.

\[65\] See, for instance, Julia Driver’s consequentialist virtue theory, which is most certainly not a virtue ethic. Julia Driver, Uneasy Virtue (New York: Cambridge University Press, 2001). Is there any reason to see her as less focused on cultivation or focused on the virtues because her view does not take them as the basic normative standard? See also the “Self Cultivation” chapter (Chapter 2) of the Mozi, which clearly discusses virtues, but is also clearly not evidence that the Mohists were virtue ethicists.
ren (benevolence), Confucius replies, “why stop at ren? This is certainly a sage (sheng)”\(^66\).

Clearly, the sage in the *Analects* is an especially high ethical ideal, as Confucius himself claims not to be a sage\(^67\) (and the Grand Minister agrees)\(^68\), and says that the sage walks the dao of the junzi every step from start to finish\(^69\), which is perhaps why the junzi is in awe of the words of the sages\(^70\). Confucius claims that he may never meet a sage, but it would be enough to meet a junzi\(^71\).

Is the presence of these “good people terms” enough to claim that the *Analects* presents a virtue ethical theory based in being sage-like? I think not. First off, noting the existence of a character ideal, even if it is clearly held in high esteem, does little to support the view that the character ideal is the central basis of a normative view. This is because we can talk about “moral saints” without needing to claim that our moral judgments are somehow based in an ideal of sainthood or the characteristics of a saint,\(^72\) and because it is possible to unpack character ideals like saints or sages in terms of conformance to some other moral standard.\(^73\) In light of this, we would need more to be able to say that the notion of the sage serves as a standard in a normative theory in the text. The evidence for

\(^66\) *Analects* 6.30. As I will argue further down, this comparison undermines claims that ren is a sort of foundational virtue standard in the *Analects*.

\(^67\) *Analects* 7.34.

\(^68\) *Analects* 9.6.

\(^69\) *Analects* 19.12.

\(^70\) *Analects* 16.8.

\(^71\) Cf. *Analects* 14.5, in which he does label Nanrong Kuo a junzi on the basis of his rhetorical question about why Yu and Ji, through farming, came to rule the world when others who cultivated physical strengths met untimely ends.

\(^72\) See Susan Wolf, "Moral Saints," *The Journal of Philosophy* 79, no. 8 (1982): 419-439. Also, notice that the Mohists similarly held the sage kings up as an ethical ideal with some authority as a model, but the Mohists are most definitely not virtue ethicists. See *Mozi*, Chapter 4.

\(^73\) Christians, for instance, can hold up Jesus as a perfect paragon of virtue without taking the Christian normative theory to be based in being Jesus-like (as opposed to being based in following God’s commands, etc.).
this further step is lacking; not enough discussion takes place in the *Analects* regarding the notion of a sage and its relationship to other moral terms to warrant a claim that it is a sort of central standard in the *Analects*. “Be sagelike” is surely a prescription that the subjects and authors of the *Analects* would endorse, but this hardly gives them a sage-based theory.

What about the notion of the *junzi*? Clearly, it is a more central topic of discussion than the sage, it has connections to morally relevant Confucian terms like “de”[^74] and *ren*[^75], and it is an esteemed character ideal contrasted with the “petty person”[^76]. It is fairly clear from the corpus that “*junzi*” was once just a term for the class of nobles[^77], so the connotation of moral excellence is likely an aspect of the term that Confucius was either adding or making more conspicuous[^78]. Was this reframing of the term, though, in service of producing a character ideal around which to center a systematized virtue ethical theory in the *Analects*? I think in this case, too, the answer has to be no.

This is because in the text, a *junzi* is often treated as somebody who fulfills some collection of moral standards that appear to be independent of one’s status as a *junzi*, and that is thus akin to a notion of moral sainthood to which people could aspire regardless of how or whether they have systematized their ethics. We are told that the *junzi* focuses on what is *yi*, but we are never told that what is *yi* is somehow determined by or based in being a *junzi*. What it means to be a *junzi* appears to be in large part identified by the degree to

[^74]: *Analects* 4.11.
[^75]: *Analects* 4.5.
[^76]: *Analects* 2.14, 4.11, 4.16, 6.13, 7.37, 12.16, 12.19, 13.23.
[^77]: See, for instance, *Analects* 11.1, and the Mozi passage cited in footnote 3 of this chapter, which clearly uses “*junzi*” to pick out a social class rather than a moral class.
[^78]: This happens regularly in moral discourse. People add moral connotations to classes as ways of telling people that they have reasons to act thus and so. Telling somebody to be “a grown up” or “a true American”, etc., lets somebody know that there are moral expectations. Confucius seems to be using “*junzi*” here as a “real noble” to extol a morally appropriate existence.
which one focuses on *yi*.\(^79\) It is also identified by a grab bag of other qualities including not harboring frustration at going unrecognized\(^80\), not abandoning *ren* (benevolence)\(^81\), acting before speaking\(^82\), taking the high road\(^83\), and being steadfast in the face of adversity\(^84\), being easy to serve but difficult to please\(^85\). These are all admirable, but we could come up with a similar list by asking people without a worked-out normative theory to tell us everything they can about what a moral saint does. Even more telling, the *junzi* does fail to be *ren* sometimes.\(^86\) In this, we can see that “*ren*” and “*junzi*” are not simply defined in terms of each other. It seems more that the *junzi* is capable of having and/or acting on *ren*, focusing on *yi*, and doing other various noble things that bear no indication of being defined in terms of something like “what the *junzi* would do”.

This still leaves the possibility that the *Analects* contains a virtue-ethical normative theory not in light of holding to a particular character ideal, but in light of taking normative judgments to bottom out in judgments about some (set of) character trait(s). The possibility that the *Analects* contains a virtue-ethical normative theory owes much to the prominence of the term “*ren*” (“benevolence”) in the text. “*Ren*” occurs 105 times throughout, in contexts that make clear that it is a moral-laden term, and it often appears as the subject of

\(^{79}\) Which makes sense if *yi* is simply a term with a similar function to the English phrase “morally right”, and a “real noble” or *junzi* focuses on what’s morally right.

\(^{80}\) *Analects* 1.1.

\(^{81}\) *Analects* 4.5.

\(^{82}\) *Analects* 4.24.

\(^{83}\) *Analects* 14.23.

\(^{84}\) *Analects* 15.7.

\(^{85}\) *Analects* 13.25.

\(^{86}\) *Analects* 14.6.
clarificatory questions and explicatory statements.\footnote{Roger Ames argues that this is because Confucius is redefining “ren,” taking an older, obscure term and imbuing it with a new meaning that his protégés are still attempting to grasp. See Roger T. Ames, \textit{Confucian Role Ethics: A Vocabulary} (Honolulu: Hawaii University Press, 2010), p. 176. My discussion of “junzi” above has certain affinities to this general interpretive line.} It is perfectly reasonable to say that Confucius was at pains to elaborate what \textit{ren} was without also saying that it plays the role of a theory-centering virtue, though, so we need further evidence to justify that latter claim.

One could give that further evidence by showing that a plausible interpretation can be constructed on which the other normative terms and judgments in the \textit{Analects} are consistently based in \textit{ren} and not the other way around. Whether we can show this or not has already been discussed in the literature in debates about the relationship between \textit{ren} and \textit{li} (ritual). If having \textit{ren} is the central, systematizing moral goal of the \textit{Analects}, then one interpretation that we could give of the relationship to \textit{li} might be what Kwong-Loi Shun has called the instrumentalist reading,\footnote{See Kwong-Loi Shun, “Ren and Li in the Analects,” in \textit{Confucius and the Analects: New Essays}, ed. Bryan Van Norden, 53-72 (2002).} on which \textit{ren} is the basis of the \textit{Analects’} normative theory and the \textit{li} are a means toward building or expressing \textit{ren}. Shun, I think rightly, sees this particular reading as inadequate, however, because it cannot account for Confucius’ general conservatism about the \textit{li} or for certain passages\footnote{Such as \textit{Analects} 12.1. Chenyang Li disagrees with this criticism, citing a different interpretation of 12.1, but agrees that Confucius’ conservatism about the \textit{li} speaks in some favor of a definitionalist reading. See Chenyang Li, “Li as Cultural Grammar: On the Relationship Between Ren and Li in Confucius’ Analects,” \textit{Philosophy East and West} 57, no. 3 (2007): 311-329.} that Shun sees as supporting what he calls a definitionalist reading, wherein to be “\textit{ren}” simply means to unfailingly follow the \textit{li}. Shun also, again I think rightly, criticizes the definitionalist reading for not accounting for the fact that Confucius accepts revision of the \textit{li} while \textit{ren} seems unsusceptible to similar revision.

To navigate these potential pitfalls of the instrumentalist and definitionalist readings, Shun
offers a reading on which being ren consists in performing the li just as getting married consists in having a marriage ceremony. If Shun is right, then maybe the normative theory of the Analects is based on something like having-ren-through-following-li.

Shun’s reading also faces difficulties, though, some of which are noted in the literature. Chenyang Li, for instance, has argued persuasively that Shun’s reading is undercut by passages that note the possibility that a person could follow the li without also being ren. Another major difficulty, it seems to me, comes from the passages that discuss the old Prime Minister Guanzhong. We are told definitively that Guanzhong had ren, a lofty accomplishment that, recall, Confucius did not even claim for himself. Whatever it means to be ren, Guanzhong did it, but what he did not do, according to one passage, was to understand ritual propriety. I would contend that if the subject of one of Confucius’ harshest criticisms in the Analects for failing to follow the li is the same person on whom Confucius heaps praise for being ren, then view of the relationship between ren and li cannot be captured by Shun’s constitutive reading, and so we cannot say that the Analects is pushing a cohesive normative theory based in having-ren-through-following-li.

This “Guanzhong objection” also presents further problems for the instrumentalist reading, as Guanzhong’s having ren despite not understanding the li means that the li neither developed nor gave expression to Guanzhong’s ren. We can, however, rescue a type of

---

90 See Li, “Li as Cultural Grammar”. The passages he mentions are 3.3 and 3.26. I would add that Confucius’ discussion of “sacrificing while present” further supports the point. Furthermore, I agree with a sort of instrumentalist reading, that ren is desirable and that performing the li in the right way helps to build ren, and that having more ren makes one more capable of performing the li the right way, much like how practicing music gives one musical skill, and greater musical skill makes one’s practice more. I think that this sort of reading, like Li’s, allows us to explain everything Shun seeks to explain except for Analects 12.1, which, I agree, can also be given an instrumentalist reading.

91 14.9, 14.16, and 14.7.

92 Analects 3.22.
instrumentalist reading of the relationship between ren and li from the Guanzhong objection. If we accept both that one can have ren without engaging in the li and that engaging in the li is a superlative method of cultivating ren. Such a reading can account for the texts’ moderate conservatism about the li because wantonly revising or sporadically following effective cultivation methods typically diminishes their effectiveness. It can also account for what appear to be “definitionalist” passages, because noting that the ren person is the person always following the li bears a similar important function on this reading as saying that, for instance, the best drummers are the ones who are constantly practicing.

In addition to avoiding Shun’s objections, this reading also fits well with a denial that ren serves as a theory-centering trait in the Analects, because ren need not have such a theoretical role in order to be a desirable goal of cultivation through the li, just as we can praise methods of cultivating kindness without this indicating that we hold a kindness-based normative theory. So, this casting of the instrumentalist reading might resolve the thorny interpretive issue of the relationship between ren and li, and do so in a way that also ultimately lends some support to my broader interpretation on which neither ren nor anything else serves as the basis of a normative theory in the Analects.

Of course, the significance of that support, as well as the overall balance of support for or against my position, depends on how we should understand “ren” and other moral terms in the Analects beyond merely what is relevant to the ren-li discussion. I need to

---

93 Which seems to make sense of Confucius’ claim in Analects 7.30 that ren is not remote. I take this passage to indicate, along with everyone’s favorite point that “ren” is “person” + “two”, that ren is a type of pro-socialness. I read that passage as saying that wanting to be pro-social is itself pro-social.

94 Percussionists in the Legends Performing Arts Association can attest that I show a similar conservatism regarding my method for playing and cleaning music.
foreclose, for instance, the possibility that ren is a virtue, the attainment of which is
definitional of proper behavior. If ren is the basic standard of a virtue-ethical normative
theory in the text, then we should find indication that one’s being ren and one’s acting in a
morally correct manner should be one and the same. When we look to the text, we do see in
the Analects the claim that if one’s will is set on ren, one could do no wrong, so there is
something in the text on which to build a case. One problem with this case, though, is that
Confucius’ mini-biography portrays him as able to give his heartmind free reign without
overstepping the line, something that can be taken as indicating that he had reached a state
where he could do no wrong. Eirik Lang Harris is typical in taking this passage to indicate
an account of Confucius having reached a state that “would ensure that he was acting in a
virtuous manner.” I agree with Harris’ reading of that passage, but think we need to square
it with the text also presenting Confucius’ claim that he is not ren. These two passages
collectively indicate that being able to give one’s heartmind free reign without overstepping
the line, i.e. consistently acting in a moral way, is something that a person can do without
being ren. If this is the case, then rightness cannot be defined in terms of acting from ren the
way rightness is defined in terms of a core virtue in a virtue-ethical normative theory.

In other passages, ren is certainly treated like a virtue but is simply one alongside
others like wisdom and courage, and is downgraded even further by a passage citing that

---

95 Analects 4.4.
96 Analects 2.4.
98 Analects 7.34.
99 Analects 9.29.
virtue triad as merely a part of the way of the junzi.\textsuperscript{100} Coincidentally, this passage also has Confucius claiming that he does not find these three in himself, which lends further evidence to both the claim that Confucius’ virtuous end state is not one that can be cashed out in terms of his having acquired ren and the claim that Confucius’ free-roaming heartmind is not the defining mark of a junzi.\textsuperscript{101}

So, there are serious problems for a reading of the Analects as containing a ren-based virtue-ethical theory. This still leaves open the possibility that the Analects contains a virtue ethical theory that, while not being specifically ren-based, still finds the locus of normative rightness in some broader set character traits that a person can embody.\textsuperscript{102} On such a reading, ren is still one virtue among many, and having some collection of virtues is what one must do in order to meet the normative standard. This approach seems even less likely to me to succeed than a ren-based account, because there is no real indication either that the virtues in the Analects are unified or that there is some privileged list of them that we definitely ought to have. There is, furthermore, no indication that only virtues in the text are accorded what looks like basic normative significance. Sometimes rules are simply presented as though they are of independent normative import.\textsuperscript{103} Sometimes, normative judgments

\textsuperscript{100} Analects 14.28.

\textsuperscript{101} Further evidence for this split between being a junzi and achieving what Confucius achieved comes from Analects 7.33, where Confucius says that he has accomplished little in the way of living like a junzi.

\textsuperscript{102} Note that one need not subscribe to a “unity of the virtues” thesis or the idea of a “master virtue” in order to have a viable virtue-ethical normative theory.

\textsuperscript{103} When Confucius tells Zigong that the saying “do not impose upon others what you yourself do not want” in Analects, it is presented as a saying that one can act on, not necessarily as the expression of some trait.
are made apparently without any needed reference to virtue terms.\textsuperscript{104} There is no positive reason to read these rules or judgments as somehow parasitic on an account of the virtues.

Could we say that instead of some collection of named virtues, Confucius’s state of cultivated spontaneity is the true goal in the \textit{Analects}? I think we can,\textsuperscript{105} but this remaining possibility does not mean that the \textit{Analects} present us with a virtue-ethical theory. I have already argued that people can want to cultivate that kind of moral mastery regardless of what their normative commitments are, and regardless of whether their commitments are systematized or not.\textsuperscript{106} I would further argue that people can even take an interest in developing a virtue concept, again, without doing so in service of what we could call a virtue-ethical normative theory.\textsuperscript{107} An interest in virtues can easily coexist with interests in other types of moral commitments like duties to family or some formulation of the golden rule, it often does in discussions among people with pre-theoretical normative commitments but no normative theory, and it appears to do so in the \textit{Analects}.

\textbf{3.2.4 A Role-ethical Theory in the \textit{Analects}?}

Another sort of attempt at interpreting the \textit{Analects} as containing a normative theory that warrants discussion here is the possibility that it contains a role-based normative theory. It warrants discussion in light of the text’s portrayal of Confucius as strongly focused on

\textsuperscript{104} For instance, in \textit{Analects} 13.18, there seems to be a simple claim about the correctness of the act of turning one’s father in. When Confucius tells Zaiwo not to criticize what has come in the past in 3.21, it seems to be a similarly straightforward imperative.

\textsuperscript{105} Certainly, this type of free roaming without overstepping the line is, even on my cultivate-first-and-theorize-later reading, central to the \textit{Analects}.

\textsuperscript{106} Recall that although ethical attainment ideals cannot exist absent normative positions, they can easily exist separate from normative theories or unified ethical views.

\textsuperscript{107} An analysis of “kindness” does not a kindness-based normative theory make.
social roles, their attendant ethical commitments, and people’s ability to live up to them.

Confucius insists, for example, that one should cover for one’s father even in the event that one’s father steals a sheep; he also clearly values being filial and respecting his elders and his lord. Roles are thus, at the very least, part of what Confucius sees as normatively relevant. How might one, though, elevate this claim to the claim that the Analects present a role-based normative theory?

On an account of role ethics that comes from Christine Swanton, role ethics “is concerned with success in meeting the point and function of a role.” Working with such an understanding of role ethics, to claim that the Analects contains a role-ethical normative theory would be to claim that the moral discussions in the Analects generally ground out at talk of roles. If this interpretation is right, then we should expect to find evidence that the normative judgments made in the Analects are ultimately indexed to particular roles and that the role-to-judgment indexing is complex enough to warrant doing it in the first place. We might, perhaps, also expect to find some positive indication that the normative status of a role is taken to be source of the valence of moral judgments of acts and people as they show up in the text.

108 Analects 13.18.
109 For instance, see Analects 2.5-2.8, wherein several students ask Confucius about being filial.
110 Analects 10.13.
111 Analects 10.3, and 10.18-10.20.
113 We do, of course see a couple of instances of this conceptual dynamic in the text. When discussing the morality of sheep stealing, the role of a son involves not turning in one’s father when one’s father steals a sheep; when discussing serving under ruler, the role of a ruler’s servant involves not decorating like a ruler. See Analects 3.22 and 13.18.
114 We would not want to call something a role ethics if it essentially boiled down to something like “Rulers: Maximize utility; Subjects: Maximize utility; Parents: Maximize utility, etc.”, even if discussions of roles were prominent.
We do not, however, find such things when we examine the text. We find rather that judgments appear with little indication that they are tied to some role, such as when Confucius says plainly that he doubts that a person who lacks honesty is acceptable.\(^{115}\) We also find talk of roles in which the role appears to be defined in terms of some degree of moral success rather than being the determinant of what counts as moral success. The discussions of the role of junzi, for instance, show its relationship to yi to be that a junzi constantly focuses on what is yi (moral)\(^ {116}\), so “junzi” appears to refer to people who diligently fulfill moral commitments, period, and not to some role that functions so as to center or ground what it would be yi for a certain person to do.

In addition to the existence of normative judgments that do not seem based in roles, and roles that do not seem to serve as a basis for some set of normative judgments, we also find a passage that I think brings into question the discreteness of roles as the Analects views them. In this passage, Sima Niu laments that he is the only one without brothers, and Zixia tells him that the junzi has everyone in the world as a brother.\(^ {117}\) Now, if “brother” (xiong and di both appear in the passage) were meant to be one of the basic, norm-determining roles in the Analects, it seems contrary to the spirit of a role-ethical theory to insist that acting toward people in one capacity (as non-brothers) is the same as acting toward them in another capacity (as brothers). We would expect the passage to say, perhaps, “worry not, you still have plenty of meaningful roles to perform properly” rather than “worry not, everyone in the world will still give you an opportunity to play the brother role if you are

\(^{115}\) Analects 2.22.  
\(^{116}\) Analects 4.16.  
\(^{117}\) Analects 12.5.
doing things right”. Presumably, if the Analects contains a normative theory that makes varied, role-based prescriptions and judgments, and the text is serious about treating “the junzi” as a term of moral accomplishment, then somebody as morally-accomplished as a junzi would not act toward everyone in the world in a manner befitting the role of a brother.

The conflict between this “brother” passage and a possible reading of the Analects as containing a role-ethical normative theory is especially stark when we look at the passage most clearly appears to advocate a system of roles. In Analects 12.11, Confucius responds to Duke Jing’s question about governance by saying “jun jun chen chen fu fu zi zi” (ruler ruler minister minister father father son son). One parsing option for this sentence clearly advocates the performance of roles: The ruler rules, the minister ministers, the father fathers, the son sons. If this is the way to read the sentence, and this is indicative of a role-ethical commitment on the part of the subjects and authors of the Analects, then it seems especially odd that Zixia would, in 12.5, blur the lines between one of these roles (jun) and another (xiong/di) that, we would presume, has similar normative qualities. Another way of parsing that phrase, (“ruler” rulers, “minister” ministers, “father” fathers, “son” sons) still involves the use of roles, but has the advantage of avoiding conflict with Zixia’s point, which it does by casting the consistent application of role-terms as a leadership strategy rather than a way of understanding people’s different normative requirements to each other.\(^\text{118}\).

Perhaps, though, the Analects contains a role ethical theory in a different sense from the Swanton account that I have been working with thus far. Roger Ames has recently argued for a role ethics reading of Confucianism that is quite different from a

\(^\text{118}\) As I will discuss in Chapter 4, the plausibility of the second parsing can lend support to my “cultivation-first” interpretation of the significance of roles in the Analects.
straightforward view of Confucianism locating the source of normativity in people’s social roles. Ames’ particular view is located in a broader picture of Chinese thought as imbued with a correlative cosmology\(^\text{119}\) in which “All activity occurs within a context and is thus collateral in nature”.\(^\text{120}\) On such a view, “no-thing or no-body has an *essence*, but can be defined only ‘correlationally’” at any given time, with differing relations holding at other times.\(^\text{121}\) This means that I cannot have virtue *qua* what I am separate from my relations (i.e. a human being), as there is no me separate from my relations. Instead, what I can have is a virtuosity\(^\text{122}\) that comes from engaging a correlative process of moving through my different relations (i.e. a human becoming).\(^\text{123}\)

Situated in this sort of overall view, Ames’ role ethics, unlike a Swanton-style role ethics that countenances things like “role duties”\(^\text{124}\), does not seek to give us any once-and-for-all prescriptions that happen to be based in roles. In fact, Confucian role ethics is, to Ames, “not an abstract theory that provides principled moral judgments… nor does it give primacy to developing a deliberate, rational means to achieve some moral end”\(^\text{125}\). It is, on some level, the only way to think about ethics for people who live with such a radically correlative and thus particularized worldview, which is why Ames treats role ethics as a “new and compelling vision of the ethical life”\(^\text{126}\) from a culture who systematically avoids making

\(^{120}\) Ames, *Confucian Role Ethic*, p. 72.
\(^{121}\) Ibid., p. 24.
\(^{122}\) Ibid., p. 87.
\(^{123}\) Ibid., p. 159.
\(^{124}\) Swanton, “Virtue Ethics, Role Ethics, and Right Action”, p.15.
\(^{125}\) Ames, *Confucian Role Ethics*, p. 164.
\(^{126}\) Ibid., p. xvii.
the “philosophical fallacy”, rather than a competing normative theory, capable of being brought into the kinds of debates typical in Western philosophy.

Ames’ view, then, appears to agree with me by default on my main point that the Analects contains no normative theory. Ames’ view does, however, still warrant discussion here, as his view of the Chinese philosophical dialectic as a whole competes with the positive view of the dialectic that I am ultimately concerned to support, namely that the Confucians of the Analects got the conversation started by discussing how to master their moral commitments, and the attempt to find standards emerged after that. On my view of the period, classical Chinese thinkers can and do make sense of the difference between a project of finding grounding normative standards and a project of developing oneself such that one consistently follows such standards. On Ames’ view, classical Chinese thinkers can make no such distinction.

The argument in favor of Ames’ position has some weaknesses, however, especially insofar as we take it to imply an argument for a way of reading the Analects in particular. First, the argument relies on the wrong sorts of evidence to mount a challenge to the position that I am advancing. Much of it relies on attempting to discuss the Confucian tradition as a connected, trans-millenial whole, which includes neo-Confucian theories. It also relies on discussing the evolution of Chinese thought and language in general, as though the Confucian view flows in perfect concert with what it means to be Chinese and to think in characteristically Chinese ways. This, however, leads Ames to focus little on accounting for other trends in classical Chinese philosophy that appear distinctly anti-Confucian, as

---

127 Ibid., p. 16.
though either they did not happen, or as though they can be easily and summarily situated in the same Chinese-as-Confucian-as-Chinese worldview.

The issue that I am raising is rather glaring in a particular passage of Ames’ book, wherein Ames quotes a paper he wrote with Henry Rosemont in which they move from saying that “to search for a universal principle in Confucianism is to square the circle, for Confucianism is paradigmatically particularistic”128 to saying that “to seek a principled ground for moral judgments in the classical Chinese texts is, to our minds, like expecting a Kantian to take specific cultural differences into account as conditions that qualify his Categorical Imperative”.129 The problem with such a move between “Confucianism is way X” and “All of Chinese thought is way X” is that it gives us no way to understand the Mohists, with their explicit call for standards that will help us determine how to make shi/fei judgments. Perhaps this view of the Mohists reflects a mere appearance that can be explained away so that we can also see the Mohists as particularists and thus properly situated in the Chinese worldview. That seems highly unlikely given the arguments in the Mozi, unless a call to move from 100 yis to one yi can plausibly be read as something other than a call for a unified morality, and unless compasses, plumb-lines, and t-squares yield situation-relative standards. It is also hard to see what, on Ames’ story, the Mohists could have even differed with the Confucians over if particularism is somehow built into the Chinese worldview; nothing motivates their dispute in the first place if, as correlative thinkers, the Mohists were not going to engage in arguments about how to unify and justify

129 Ibid., quoted in Ames, Confucian Role Ethics, p. 165. The emphasis is mine.
moral judgments. If the Mohists cannot be brought into the fold, and they really did disagree with the Confucians, then we must either conclude that they were not really Chinese thinkers, which is obviously a non-starter, or we must conclude that Ames’ particular conception of the Chinese worldview gives us an inappropriate set of background assumptions to bring to bear on an interpretation of the warring-states texts, including the *Analects*.\(^{130}\)

There is nothing wrong with trying to find a sort of reflective equilibrium between a mutually reinforcing holistic interpretive theory of some texts and a theory of the intellectual and linguistic background that informed those texts. This is, in fact, what all interpreters need to do to some degree. The mistake is in treating Confucianism as the primary representative of Chinese thought, and taking the best background theory to be the one that best explains all and only the diachronic Confucian corpus. There is simply far less reason for taking whatever worldview we could simultaneously attribute to Confucius and Zhu Xi as evidence of the content of the *Analects* than there is for taking whatever worldview we could simultaneously attribute to Confucius and Mozi as evidence of the content of the *Analects*. As such, I do not see Ames’ role ethics interpretation as a plausible option.

3.2.5 An Exemplarist Normative Theory in the *Analects*?

The final possible way of reading the *Analects* as containing a normative theory that I would like to discuss here is Amy Olberding’s reading of the *Analects* as containing an

---

\(^{130}\) I would add that those background assumptions seem uncannily like they are tailor-made to make Confucian ethics look superior to pre-Qin rivals by finding a way to interpret it that makes it, essentially, the only game in town.
“exemplarist moral theory”\textsuperscript{131}. Olberding describes her interpretive project as attempting to “discern a governing logic that renders the Analects’ compelling moral sensibility intelligible as moral theory”, \textsuperscript{132} using an interpretive approach in which she “[hews] narrowly to the text, largely forgoing the supplements of later Confucian tradition.”\textsuperscript{133} Taking this approach, she argues that the best explanation of the Analects is that the subjects and authors were developing an exemplarist moral theory, i.e. a theory on which normative standards are grounded through experiences of moral exemplars to which we can refer directly, and on which normative theorizing is a process of discussing exemplars so as to unpack the totality of their example into a manageable moral vocabulary so as to systematize our normative judgments. This explanation fits with an origin myth she presents “in which Confucius and the authors of the Analects are moved to their philosophical activity by their admiration for exemplars.”\textsuperscript{134}

Olberding’s view has a number of important strengths that allow it to avoid many of the criticisms that I have made of other views in sections 3.2.2-3.2.4. For instance, her interpretive project is calibrated such that it avoids any of the issues I raised for theories justified by reading later Confucian views into the Analects\textsuperscript{135}. The theory avoids proliferating abductive inferences about what kinds of views the subjects and authors of the Analects must

\textsuperscript{132} Olberding, \textit{Moral Exemplars}, p. 1.
\textsuperscript{133} Ibid, p. 2.
\textsuperscript{134} Ibid., p. 14.
\textsuperscript{135} The content of later theories in a dialectic can certainly be brought to bear as evidence in the interpretation of earlier texts, because we do need some story of how the dialectic got from text A to text B, and our ability to tell that story well speaks in favor of our interpretations of each text. This gives us a far different than we get if we simply try to make text A cohere with text B, one that, given what we know about how dialectics tend to operate over centuries, surely gives us an inaccurate reading of text A.
have held on fundamental but undiscussed elements of their moral theories. It does this by simply explaining away what appear to be lacunae in the *Analects*’ normative discussions, arguing that the *Analects* offers “a style of moral theory that successfully circumvents some of our typical expectations of moral theory.” It also accounts for many of the noticeable inconsistencies in the moral discussions in the text in a way that allows them to exist as a natural side effect of an exemplarist project in motion rather than justifying them, as Ames does, as a consequence of a metaphysical/linguistic orientation that precludes generalizing thoughts and thus leads inevitably to a strongly particularistic moral understanding.

I will argue, though, that despite, and in part because of, the strengths of Olberding’s approach, we still have reason to prefer an interpretation on which the *Analects* simply does not contain a normative theory. What I say here will not, of course, be the final word, because the final word depends in part on the complex process of debating whether the interpretation I give in Chapter 4 is a more satisfying, elegant, and complete account of the text than Olberding’s account. I intend to establish here simply that an account of the *Analects* on which the subjects and authors are concerned to cultivate compliance to their pre-theoretical moral judgments (rather than to construct a normative theory) shares much of what speaks in favor of Olberding’s exemplarist account over rival normative theory interpretations, and does so while avoiding a few of the more serious objections to Olberding’s approach.

An interpretation on which the *Analects* contains no normative theory, for instance, shares with Olberding’s view that it requires no justifications drawn from overestimating the

---

evidentiary value of Neo-Confucian views for the interpretive task at hand. It also allows one, in similar ways, to avoid proliferating inferences about what answers the subjects and authors must have held on matters typical to moral theory. Occam smiles when we can answer the question of “what in the text can serve as a foundation for moral theory”\textsuperscript{137} without invoking a largely extra-textual account of what theory of human nature or flourishing the subjects and authors of the \textit{Analects} must have held to ground their moral theory, which Olberding accomplishes by pointing at exemplars in the text, and I accomplish by simply answering that there is not yet a moral theory to ground, though there are, of course, moral views, many of which, as Olberding notes, are discussed and explained with reference to particular people. My interpretive view also shares the virtue of accounting for the inconsistencies in the text’s moral discussions, and does so by simply saying that we should expect to see such a thing in a discussion among people who have not yet gone through the process of working their way from their views to a normative theory that can then be used to bring those views into greater consistency.

In addition to accounting for the text with little recourse to extra-textual speculation, Olberding’s interpretative theory is substantively strong in that it makes central to the normatively theory something that is extremely prominent in the text. Discussion of moral exemplars occurs often throughout in the \textit{Analects}, as do rather vivid accounts of the subjects of the \textit{Analects} that Olberding argues we are offered for contemplation in an exemplarist spirit.\textsuperscript{138} Olberding’s ability to account for the “narrative” elements in the text also lends credibility to her view, as she is able to account for parts of the text that many philosophical

\textsuperscript{137} Ibid., p.45.
\textsuperscript{138} Ibid., p. 105.
interpretations have passed over due to “an ‘embarrassment’ about at least some of the text’s narrative detail, with their curiously close attention to apparently trivial detail.”

The view that cultivation rather than normative theory is the focus of the *Analects* shares these substantive strengths, placing talk of personal cultivation, long recognized to be a central aspect of the text, at the heart of the text while accounting naturally for the various types of moral talk in the text. It also has the virtue of finding an important role for many of the narrative elements Olberding is concerned not to abandon. Discussing moral exemplars in narrative details is an important way of setting out and justifying a method for becoming one.

Having shown that my reading shares many of the strengths of Olberding’s reading, I will attempt, in the criticisms that follow, to show that a cultivation-focused reading of the *Analects* is preferable to an exemplarist moral theory reading. I should note, here, however, that the argument between us is really one of how strongly we each see the text and the discussions within the text aiming toward certain goals given what we find in the discussions, rather than being a matter of one reading being correct to the exclusion of another, as there is a good deal of compatibility between our readings. Discussions of how to become people who do what we think we ought to do can make use of exemplars no less than discussions of which normative commitments we should take up, and conversational moves that respond to those different goals can easily flow into each other and sometimes even overlap. I would say that to whatever degree we see the project of normative theorizing getting off the ground in the *Analects*, it is, as Olberding argues, by way of discussions of exemplars and

---

their practices, though I do not see the search for answers to the normative question as motivating the discussions as strongly as the search for answers to the cultivation question.

One substantive weakness of Olberding’s account is the way she treats the prominent notion of *li* (ritual) in the text. On Olberding’s reading, we “are not urged to follow the code of *li* because moral rulers said we ought, but because they themselves enacted the *li.*”\(^{140}\) The *li* then are, on her account, a sort of distillation of the acts of exemplars. The motivation for treating the *li* in this manner is clear enough: as such a prominent, clearly moral term in the text, to not successfully account for it in exemplarist terms would leave a sizeable hole in the range of the text’s moral discussions for which Olberding’s reading can account. Unfortunately, this approach to the *li* faces some serious difficulties. One is that the *li*, as we can recall from the discussion above of Guanzhong, give different prescriptions to people in different roles. The *li* contain too many disjunctive, role-laden prescriptions to merely reflect the actions and manners of some handful of exemplars. Additionally, the *li* clearly pre-date the *Analects*’ project of spelling norms out from the manners of exemplars, so it is not clear when or how the *li* would have been developed so as to “lift from the complexity of individual exemplary people and lives the practices to which those lives point us”.\(^{141}\) If they were not developed according to the exemplarist procedure, and yet Olberding sees them as providing a significant portion of the moral content of the Confucian view in the *Analects*, then this speaks against a reading of the *Analects* as presenting a normative theory grounded in exemplars. If they were developed using an exemplarist reading, then we need to accept that the broader society was somehow

---

140 Ibid, p. 94.
141 Ibid. p. 95.
engaging in exemplarist theorizing throughout the history of the *li* up through the *Analects*, which is the kind of extra-textual architecture that Olberding’s explanation so successfully eschews in other places.

While this might undermine Olberding’s specific argument, it need not undermine the general notion that discussion of exemplars was used to shed light on the *li*, their importance, how best to perform them, and so on. It also, of course, need not undermine the idea that discussions of exemplars were a major mode of investigating the ethical in the *Analects* that helped shed further light on Confucians’ normative commitments and how they related to one another. Those can both be the case without being as neatly, systematically so or as in service of normative theoretical aims as Olberding claims. We can perhaps also find a way to rescue the idea that whatever incipient normative theorizing we find in the *Analects* happens in an exemplarist mode; if, for instance, we take a mostly practical stance toward the *li* as cultivation tools rather than seeing them as a part of the *Analects’* normative view, then the unavailability of an exemplarist explanation of their worth would still leave an exemplarist reading compatible with the idea that whatever should be taken as a normative standard in the *Analects* is susceptible to exemplarist explanation. This would, I believe, serve to make Olberding’s general account of the structure of the *Analects’* discussions and my cultivation-focused reading all the more compatible by highlighting how discussions of exemplars can shed light on moral cultivation methods.

A related objection that I have to the exemplarist theory reading results from looking at the *Analects’* robust stock of moral concepts in light of Olberding’s claim that “exemplars
enjoy both sequential and conceptual priority”. My objection is that it is difficult to see how we can give these moral concepts all an exemplarist reading. After all, comprehensive exemplars are, according to Olberding, “quite rare”\textsuperscript{143}. Such a limited range of examples could not serve to ground the rich moral vocabulary of pre-	extit{Analects} China in a way that both respects their use and protects the conceptual priority of exemplars, and there is no positive indication that part of the project of the 	extit{Analects} involves casting aside their received moral terms to shake out a new exemplar-grounded set.

We can perhaps supplement the grounding examples of total exemplars by looking at what Olberding calls “partial exemplars”\textsuperscript{144}, people that each embodies at least some worthy quality. The problem with claiming that these partial exemplars can yield an experience-base to help ground moral terms is that it takes a preexisting stock of terms to pick out the parts of the exemplars we are deeming admirable; a total exemplar provides us a clear enough “that” to fix the reference of “good person” without requiring the use of moral vocabulary, so it can play a concept-grounding role. Pointing at somebody’s exemplifying courage, or kindness, or patience, however, require a person to use concepts to distinguish that to which we are intending to refer from that to which we are not intending to refer. Maybe one could argue that the moral concepts that we use to pick out partial exemplars are ultimately grounded in total exemplars, but that does not seem like it could be the case given our need to turn to partial exemplars in the first place. So, either the experience base is too small to plausibly ground a community’s moral concepts, or we attempt to work from an experience

\textsuperscript{142} Ibid., p. 23.
\textsuperscript{143} Ibid., p. 37.
\textsuperscript{144} Ibid., p. 38.
base that requires us to utilize concepts that we already understand without some grounding story. Either way, the difficulty in giving a satisfying account of how some combination of comprehensive or partial exemplars are conceptually prior to the moral concepts, especially when the *Analects* says nothing explicit about it, speaks against a reading of the text as advancing an exemplarist normative theory.

3.3 If not a Normative Theory, Then What?

By pressing the foregoing objections to even Olberding’s inventive interpretation, I hope to have completed showing rather sufficiently that the prospects for reading the *Analects* as containing a systematized normative theory are dim. While there are clearly normative views being expressed, a rich moral vocabulary, and even, if Olberding is correct, a preferred method of engaging in discussions that can lead to greater normative systematization, I think the best reading of the text is one in which the project of systematizing the Confucian normative view was still in its initial stages. On my reading, discussions making use of the moral vocabulary had not yet yielded the kind of systematized views that we would see later in Mozi, Mencius, and Xunzi, though we can see how that vocabulary was ripe for attempts at systematization, including the competing attempts of Mencius and Xunzi that each gave priority to different aspects of the moral vocabulary we see in the *Analects*. ¹⁴⁵ This is because, on my reading, the project of constructing a normative

---

¹⁴⁵ The fact of these divergent courses of development for later Confucianism, I would argue, further supports the idea that the earliest Confucians had not yet systematized their theory, leaving room for different ways of systematizing their views. The later Mohists, by contrast, were still every bit as committed to the benefit standard as the early Mohists, arguably because that standard was so clearly laid out in early Mohism.
theory was less of a focus of the Confucians than the project of constructing a method of
cultivation was.

In summary, on my account, the text is primarily focused on discussions and
descriptions of a method for cultivating consistent behavioral compliance with the set of
normative commitments that Confucius and his school had received and accepted but had
not gotten very far in systematizing. The basic contention of my interpretive argument is
that such a reading does more justice to the totality of what we find in the text than any of
the readings attempting to pull a normative theory from the text do, given that what we find
in the text includes the discussions of various virtues, statements of various moral rules,
various moral judgments, a conspicuous focus on cultivation, the numerous and prominent
discussions involving the terms “ren” (benevolence) and “li” (ritual), and given, of course,
the criticisms that I have raised of the various normative theory interpretations of the
Analects.

If I am right, and the Confucians of the Analects had some shared set of pre-
theoretical normative commitments, and were primarily concerned with developing a
method of moral cultivation that would help them consistently meet those commitments,
then we should expect to find in the text the following three things: One, we should expect
to find a focus on the project of shaping a person in morally-relevant ways.146 Two, we
should expect to find evidence in the text that the Confucians of the Analects shared some
set of pre-theoretical normative commitments such that it would make sense that they were

146 And I mean this in a de dicto sense rather than a de re sense. In other words, it would not suffice, in arguing
that the Analects does not present a theory of moral cultivation, to argue that the true morality is utilitarianism,
and that since the Confucians are not attempting to construct a method for increasing one’s capacity to
maximize utility, they are not really concerned with changing themselves in morally-relevant ways.
collectively aiming to cultivate compliance to them. Three, we should expect to find an
account of the methods by which this moral shaping can take place.

It is not difficult to find passages displaying a focus on shaping people in morally
relevant ways. Confucius’ brief autobiography, leading from his setting his heart-mind on
learning at age fifteen and ending with his being able to give his heart-mind free reign
without overstepping a line at age seventy,\textsuperscript{147} clearly supports the idea that he advocated
putting in effort toward cultivating oneself for a specific end. It is clear that his students are
in for a similarly transformative education,\textsuperscript{148} whether or not that’s what they wanted when
they signed up.\textsuperscript{149} They are there to be sharpened like a tradesman’s tool, and part of that
sharpening comes from engaging with people who exhibit pro-sociality.\textsuperscript{150}

Establishing the second point, that the Confucians of the \textit{Analects} appear to share
some set of pre-theoretical normative commitments, requires not only that I present
evidence that the Confucians in the text behave in their discussions similarly to people with a
shared set of moral commitments that do not appear to have been systematized, but also
that I can give an answer to the question of what motivated all of those discussions in which
morality was a focus if it was not primarily the attempt to systematize or justify them. My
answer to this question, of course, is that they were largely discussing how to master the
moral commitments that they already had. This goal clearly seems important enough to be

\textsuperscript{147} \textit{Analects} 2.4.
\textsuperscript{148} See \textit{Analects} 2.12. Confucius does not want to merely train functionaries to perform official state rituals. He
wants to train people. In \textit{Analects} 5:10, Confucius uses the language of carving and building in explaining why
upbraiding Zaiwo would be pointless. In 5.8, students ask about whether other students have \textit{ren}. 9.22 and 1.2
both give plant analogies describing proper growth. Examples abound.
\textsuperscript{149} \textit{Analects} 2.18 records Zizhang’s intention to study as a means toward taking an official office. In 8.12,
Confucius complains that his students let thoughts of an official salary interrupt their cultivation.
\textsuperscript{150} \textit{Analects} 15.10.
motivating, could easily be taken up by people with a shared set of pre-theoretical normative commitments, and would likely necessitate a good deal of discussion involving moral terms like “ren” and “yi” that we find in the text. So, this is at least a coherent enough picture to be a plausible. Establishing that this plausible picture actually fits the text, though, requires me both to spell out what it looks like for people’s discussions to display pre-theoretical normative commitments, and to show that this is in fact how the discussions from the text look.

I consider our pre-theoretical normative commitments to be the set of normative commitments that we all have in virtue of living our lives and learning to make moral judgments, separate from the completion of an attempt to systematize them, give parsimonious accounts of what gives them their correctness, or otherwise give a grounding justification for them. Pictures of our pre-theoretical moral judgments abound in the moral psychology literature. Many studies show that people generally tend to moralize in a number of ways; for instance, that we sometimes take broadly utilitarian concerns as most important and sometimes take broadly deontic concerns as most important is part of the basis of Joshua Greene’s dual process model of moral cognition. We also, obviously, use virtue

---

151 You may want to ask me at this point, “Doesn’t this answer rest on assuming the correctness of your moral enhancement account of the rest of the text?” I would answer that it does to some degree, but not in any pernicious way. I have a bevy of independent arguments for the correctness of that moral cultivation account just around the corner, so while different elements of my interpretation do mutually reinforce one another, that reinforcement is not nearly so direct as to be circular. Furthermore, any interpretive argument that hinges on the overall coherence of an account of the text, as mine does, will require that we take the correctness of one element of the overall account as evidence for the correctness of another element of the overall account, and vice versa.

terms, from telling people to be nice, to talking about honesty, courage, and kindness. Sometimes, people make judgments without knowing quite why they are making those judgments, which is obviously evidence of their lacking a systematized normative theory to support those judgments.\textsuperscript{153} Pre-theoretical moral judgments are often consistent to a certain degree and can easily be considered competent judgments by one’s peers, but they are not yet organized under a unifying explanation.

Now, if we turn to the \textit{Analects}, we can see a moral discussion that appears to have similar elements. Some discussions are couched in terms that look like rule-following.\textsuperscript{154} Some passages appear to discuss admirable traits of a person.\textsuperscript{155} There appears to be a notable virtue, \textit{ren}, but this virtue is overshadowed (normatively, not in terms of textual prominence) by the character ideals of the sage\textsuperscript{156} and the efficacious person (\textit{shanren}善人), which are only rarely talked about in comparison with the character ideal of the \textit{junzi} (君子), although both are claimed to be higher and more rare than a \textit{junzi}.\textsuperscript{157} There are passages in which students are told to focus on what is moral, without reference to a standard for determining that, such as where Confucius says that the \textit{junzi} always goes with what is \textit{yi}.\textsuperscript{158} Together, this all indicates that the Confucians had a robust moral vocabulary, they were

\begin{footnotesize}
\footnotesize
\textsuperscript{153} See, for instance, Haidt, “The Emotional Dog”.
\textsuperscript{154} \textit{Shu} appears pretty straightforwardly to be a rule for conduct. See \textit{Analects} 15.24. The Chinese has Zigong asking if there is a \textit{yan} (expression) that is \textit{ke} (permissible) to act on for the rest of one’s days. If that is not a description of a rule, I don’t know what is. See also 5:12, wherein Zigong claims to be in a place where we would naturally follow such a rule. Also, see 3.21 for Confucius giving Zaiwo some general behavioral commands.
\textsuperscript{155} For instance, \textit{Analects} 2.5-2.8 have students asking Confucius more about what it means to be filial.
\textsuperscript{156} See \textit{Analects} 6.30.
\textsuperscript{157} Ibid.
\textsuperscript{158} \textit{Analects} 4.10.
\end{footnotesize}
concerned with being moral, and, like most cultural peers who decide to live together, they largely agreed in their moral judgments. This is all something that people can do without having completed the formulation of a general normative theory, especially one specified enough to let us clearly discern whether it is a virtue ethic, a role ethic, a rule-based ethic, etc. Considering this, I think that rather than trying to find a normative theory in the text, and attempting to dispel all of the textual tensions with which even the best candidate normative theory interpretations of the *Analects* have to grapple, it is plausible to claim that those tensions were there in the original discussions, because the Confucians had not yet worked out a systematized account of their moral standards, having concerned themselves more immediately with cultivating the ability to follow their standards.

The third point, that the text prominently contains an account of how morally-relevant shaping can take place, requires far more argument to establish; it essentially involves constructing an interpretive theory of the *Analects*. As such, in Chapter 4, I lay out the positive side of my interpretive argument, wherein I argue that an account of moral cultivation is central to the *Analects*, and I lay out in detail what that account is.
Chapter 4: Moral Cultivation in the *Analects*, Part 2: The Method of Moral Cultivation in the *Analects*

4.1 Introduction

Having presented the critical portion of my interpretive argument in Chapter 3, i.e., that we should not interpret the *Analects* as presenting a systematized normative theory,¹ I will now present, in this chapter, the positive side of my interpretive argument. A major part of this positive interpretive argument, of course, involves simply giving a suitably detailed account of my interpretation so as to display its virtues. The discussion is split into two main sections. In 4.2, I discuss the prominent notion of ritual (li) and explain what I take to be the various aims of performing rituals given the discussions of them in the *Analects*. In 4.3, I discuss social role terms and their use in the *Analects*’ method of cultivating moral behavior.

¹ As I make clear in Chapter 3, this is not to say that the Confucians in the *Analects* do not have any sort of first order normative views, or even that there is no evidence in the *Analects* of the Confucians doing what could be called first-order normative ethical philosophy, working out both how to make their accepted set of moral commitments more consistent and how to justify them. It is only to say that the *Analects* does not contain a systematized normative theory, which I take to be a completed attempt to systematize one’s normative judgments that provides an explanation (which may be more or less successful) of why they are correct (in the broad sense of “correct” that transcends realist/anti-realist boundaries), either by bringing them into consistency with some general normative standard (be it rule-based, characterological, metaphysical, or otherwise) or by giving some account of normative correctness on which some set of seemingly inconsistent judgments all deserve to be correct (which allows for more particularist normative theories but still requires some statement and defense of that particularism).
4.2 Rituals (li 禮) and Their Aims

In this section, I aim to show a detailed picture of how rituals serve as vehicles for moral cultivation in the *Analects*. We should note, initially, that there is some good evidence from the texts in favor of thinking that the *li* were meant to play such a role. One passage, for instance, appears to show that engaging in the *li* is part of the process of developing one’s *yi* (morality):

The Master said, “Having *yi* as one’s basic disposition, developing it in *li*, expressing it with modesty, and completing it with making good on one’s word: this then is a *junzi*”.

Here we see a person who has some moral commitments and a motivation to follow them, i.e. a basic disposition that is *yi*. The passage tells us that this basic disposition is *xing* 行 (which we can understand as conducting, progressing, proceeding, etc.) by way of *li*. The *li* allow one to get to a point where one is able to follow through on what one says, which we are safe to assume generally involves commitments to conduct oneself in a certain way. I believe that this way of reading this passage helps make sense of another passage in which Confucius explains that the meaning of one of the Songs is that “The application of color is to the unadorned”. Zixia asks if this means that the *li* themselves come after, a question Confucius commends as thought provoking. I think that Zixia’s question points to a similar relationship between one’s basic disposition being *yi* and one needing to train oneself with the *li* so that this disposition comes out properly. If Zixia’s question reflects what the *li*

---


3 *Analects* 3.8.
meant to the earliest Confucians, and I think the existence of this passage in the *Analects* is at least some evidence that it does so reflect the early Confucian view of *li*, then the *li* here can be seen as a moral training tool.

On my reading, engaging in the rituals as Confucius instructs is intended to create certain changes in a person, but what changes specifically? What effects is this Confucian moral implementation meant to have on a person such that they are able to more consistently accord with a morality? First, they instill certain attentional tendencies and capacities. They train us to pay attention to the world in a way that keeps us mindful of what in our experience is morally relevant, i.e. what we need to attend to in order to do the right thing. Second, they serve to practice and thus to strengthen certain emotions, emotions that motivate behavior in ways that keep us from swerving into immorality. Third, they strengthen our self-control, or, in more modern language, they strengthen the cognitive control mechanisms that allow us a measure of control over our emotional responses and more automatic decision-making processes. Finally, they create situations in our social and overall environments that are conducive to moral behavior.

We can see how rituals are designed to fulfill the first function of training our attention by looking at the sorts of things that are typically governed by the *li*. We can see, in Book 10, which explains many of Confucius’ *li* performances, a concern for details ranging from how mats are placed⁴, to how to communicate in certain situations⁵, to one’s facial and bodily comportment⁶, to how one should respond to receiving gifts,⁷ to how one

---

⁴ *Analects* 10.12, 10.18.
⁵ *Analects* 10.1, 10.2, 10.3, 10.10, 10.17, 10.21.
⁶ *Analects* 10.2, 10.3, 10.4, 10.19, 10.24, 10.25, 10.26.
should make sacrifices \(^8\) and how one should dress \(^9\). Elsewhere, we can see that attention was paid to proper pronunciation in performing \(lù\) \(^10\) as well as in reciting the Songs and Documents.

There is something to this attention to detail that makes it a skill that lends itself to success in many endeavors. UCLA basketball coach John Wooden famously began his freshmen’s careers with a lesson on how to properly wear their socks, a lesson that set the stage for attending more assiduously to one’s duties in practices and games. A greater attention to details, period, can lead to greater attention to the important details in a given endeavor like one’s ethical life. Something that comes through strikingly in Book 10 is the level of attention to detail that Confucius is able to maintain throughout his daily life, and I think it is plausible to say that the purpose of this look into this aspect of Confucius’ life is to give us a picture of the mindfulness that it takes for one to reach Confucius’ level of cultivation.

Every ritual performance is, in some respect, an attentional task, just like the performance of a dance, a play, or a piece of music. One must prepare with an eye to detail (are the mats exactly where they need to be?), be aware of oneself throughout the performance (Am I bowing deeply enough? Is my face solemn enough?), be aware of the details of one’s surroundings (Has the guest left the hall yet? Is the person in front of me wearing mourning clothes?), and pay attention to all of this with an eye toward what the

---

\(^7\) Analects 10.16, 10.18, 10.23.
\(^8\) Analects, 10.11, 10.9, 10.18, 10.23. See also 16.6, which implies that one should also be aware of the countenance of others.
\(^9\) Analects 10.6, 10.7, and 10.19.
\(^10\) Analects 7.18.
rituals call for in a given situation. The ritual of asking questions upon entering the Great Ancestral Hall\textsuperscript{11} forces one to attend to everything in one’s environment, and forces those answering the questions to do likewise. These attentional tasks strengthen our ability to be mindful of the details of our environment, and this in turn helps us to attend to details that are morally relevant.

Of course, the rituals are not merely all-purpose attentional exercises. Confucius appears to place special importance on details that add to what we can call the moral character of the performance. We can see this in the passage in which Confucius expresses acceptance for one alteration in the performance of a ritual, and condemnation of different alteration:

“The use of a hemp cap is prescribed in the observance of ritual propriety. Nowadays, that a silk cap is used instead is a matter of frugality. I would follow accepted practice on this. A subject kowtowing on entering the hall is prescribed in the observance of ritual propriety. Nowadays that one kowtows only after ascending the hall is a matter of hubris. Although it goes contrary to accepted practice, I still kowtow on entering the hall.”\textsuperscript{12}

Whether the cap is made of hemp or silk is a detail that can go by the wayside for Confucius, as the difference between wearing silk and wearing hemp does nothing but make the ritual more affordable. The respect that is shown through kowtowing before entering the hall, though, is a detail of the ritual that draws our attention to relations between ourselves and the other members of our community and the ways in which our body language, movements, and decisions can be used to display respect. People deciding not to kowtow

\textsuperscript{11} See \textit{Analects} 10.21.
\textsuperscript{12} \textit{Analects} 9.3.
are failing to be deferential, and to fail to be deferential is to act on an implicit judgment that one is of a higher station than those who do defer.

This is not to say that, for instance, manners of dress prescribed in the *li* do not themselves draw our attention to morally relevant facets of our environments. The wearing of mourning clothes, for instance, serves as a constant reminder to the bereaved of the loss of one’s parents, the respect that is owed to them, and the love and care that they gave. It also serves to let others know that one is bereaved, which signals to them that there are certain ways that they should act in order to display an attitude of condolence, such as taking on a solemn appearance or leaning forward in their carriage when they pass by.\textsuperscript{13} Essentially, wearing mourning clothes helps to draw one’s attention to the proper emotional state for oneself to be in, and it also serves as a signal to others that one is in a bereaved state. For those non-mourning others, the ritual makes conspicuous an emotional state that typically engages one’s sympathies and so presents a clearer view of how one may need to be treated at the present time. This helps in turn to make others’ emotional states a more conspicuous part of our environment; it puts them “on our radar” so to speak.

We can also see this strategy of training one’s attentional tendencies and capacities in the Confucian focus on archery. Archery, one of the six arts taught in the Confucian school, was an activity specifically guided by the *li*\textsuperscript{14}. We are told that the actual shooting is bookended with the archers greeting and making way for one another, and then later drinking a salute. It has built within it certain practices of respect. It is in the shooting, though, where I believe its most powerful aspect comes into play. Confucius tells us,

\textsuperscript{13} *Analects* 10.25.
\textsuperscript{14} See *Analects* 3.7.
“Marksmanship does not lie in piercing the leather target, because the strength of archers varies.” Archery, then, is not a contest of strength, but of something else. What that something else consists of, I believe, is one’s general attentional capacities and self-control. The ability to hit the mark involves taking in one’s situation, knowing when to act (between heartbeats, so I have heard), being able to calm oneself, control one’s breathing, slow one’s heart rate, etc. We are told in the Zhongyong that archery gives us something close to the way of the junzi, because if one misses the target, one looks to oneself to figure out what went wrong, rather than attempting to find some excuse. The type of training that archery represents is one that keeps us focused intently on our situation, our response to it, and whether that response is adequate.

In addition to the rituals passed down through the Zhou that Confucius is concerned to perform correctly, we can also see in the Analects what I would like to call the “ritual strategy”, the strategy of structuring one’s world, mental patterns, etc. in ways that extend the type of training that we get from the li themselves. I have a suspicion that the students at the Confucian school, and Confucius himself, were so taken with the effectiveness of cultivation via structuring their behaviors according to the li that they began utilizing the basic structuring strategy of the li in ways that were not specifically outlined by the li. The Analects gives an example of a daily ritual that I would argue utilizes the ritual strategy:

Zengzi said: “Daily I examine my person on three counts. In my undertakings with people, have I failed to do my utmost? In my interactions with colleagues and friends, have I failed to make good on my word? In what has been passed on to me, have I failed to carry it into practice?”

15 Analects 3.16.
17 Analects 1.4.
Asking these three questions has the effect of training one’s attentional tendencies and capacities. It orients one’s day toward an ultimate, evaluative event. This serves to draw one’s attention more consistently toward one’s conduct and how it is going to be assessed at the end of the day. If I get used to knowing that my day will end with an evaluation of my conduct, the quality of my conduct will, over time, become an ever-present part of my world. Finding daily answers to those questions also refines one’s capacity to notice the morally relevant features of one’s behavior. In asking daily if one has done one’s utmost, one can become more aware of failures to do so that may have previously escaped one’s notice, and so one can be more aware when a similar situation is arising in real time. Just as remembering where in a piece of music one made a mistake will lead one to be more aware of how one is performing the same phrase the next time, noting a moral failure will draw one’s attention to a type of situation that may arise again and lead one to alter one’s approach.

We can see this strategy of training a habit of attending to one’s conduct, of shaping one’s attention so that there is a constant focus both on one’s conduct and on the morally relevant features of one’s world, being put to work, for instance, in passages in the Analects that implore us to focus on yi, especially when presented with an opportunity for gain.18 When what is yi can hold people’s attention like this in general, it is more likely to hold people’s attention in situations in which the right thing to do is at odds with an opportunity for gain, and when the right thing to do is able to hold people’s attention in these kinds of
situations, people are far more likely to successfully accord with their normative commitments in their daily conduct.

The second way that rituals help us cultivate a greater ability to meet our moral demands is by allowing us to practice and strengthen certain emotions. It is fairly easy to see how Confucian rituals work to do this. The ritual performance, in Confucius’ developmental method, can be seen as a rehearsal of an emotion that he thinks will motivate productive pro-social behavior, rather than merely being a structured interaction that we can consider to have been performed correctly as long as it contained the right moves, clothing, etc. We can see this in a number of passages in the Analects.

First, we can see it in instances in which the *li* are expected to have an effect in shaping certain emotional dispositions that, left alone, would not be productive or pro-social in the right kind of way:

The Master said, “Deference unmediated by *li* will be lethargy; caution unmediated by *li* will be timidity; boldness unmediated by *li* will be rowdiness; candor unmediated by *li* will be rudeness. Where *junzi* are earnestly committed to their parents, the people will aspire to *ren*; where they do not neglect their old friends, the people will not be indifferent to each other.”

The *li* here are said to be effective in refining one’s dispositions, which are arguably conceived of as what we today would call emotional/motivational. It is interesting that we are told, after this list of what the *li* help mediate, that the *junzi*’s effect on the quality of the people comes from their commitment to their parents and not neglecting friends. When looked at in light of Analects 2.7, in which Confucius says that care for one’s parents cannot be considered filial piety unless it is performed with respect, and Analects 2.8, in which

---

19 Analects 8.2
Confucius says that filial conduct is all about showing the proper countenance, we can see that the remark about being earnestly committed to one’s parents is a matter of caring for them while actually embodying and displaying respect. That this is mentioned right after a statement about what the li can do in shaping one’s dispositions, it seems to me, indicates that this passage informs us of the li’s role in creating the types of emotions that one needs to be a truly filial progeny.

The emotion-shaping function of the li can also be seen in the role that it plays in the political philosophy of the Analects:

The Master said: “Lead the people with administrative injunctions and keep them orderly with penal law, and they will avoid punishments but will be without a sense of shame. Lead them with excellence and keep them orderly through li, and they will develop a sense of shame, and moreover, will order themselves.”

The advantage that the li have over legalistic punishments is that, unlike punishments, which only incentivize certain behaviors (including learning how to skirt the law), the li have the function of producing emotions that lead to a greater ability to accord with the same normative injunctions that the punishments are meant to enforce. The li, in other words, get ethical results, and they get those ethical results by developing moral emotions within the people.

We can also see the emotion-shaping function of the li in how, with respect to the ritual of sacrificing to one’s ancestors, Confucius puts a new spin on an old guideline for conducting a proper sacrifice:

---

20 Analects 2.3.
The expression “sacrifice as though present” is taken to mean “sacrifice to the spirits as though the spirits are present”. But the Master said, “If I myself do not participate in the sacrifice, it is as though I have not sacrificed at all”.\(^{21}\)

The presence of the spirits, was, to Confucius, not really a primary concern. We are told in the *Analects* that Confucius did not focus his energies on discussing spirits.\(^{22}\) I think, accordingly, that we should read Confucius’ reframing here of the saying about sacrifices in light of this lack of significant concern with the spirits. In this light, it makes sense to see the point of the sacrifice here as what the performer of the will get out of it, especially when the performer is genuinely present in the sacrifice. What the performer gets out of it, it seems rather obvious, is the experience of sacrificing, of being fine with having less than one otherwise could have had rather than feeling anxious over the loss, being grateful for what one has rather than being covetous of more, and of placing respect and gratitude above one’s cravings; one’s genuine presence in experiences like these make them more salient and affecting as forms of practice. This is perhaps why we are told in Book 10 that “Even with a simple meal of course grains and vegetable gruel, [Confucius] invariably made an offering, and did so with solemnity”.\(^{23}\) Every meal is an opportunity to practice the enactment of these pro-social feelings and coping skills to the exclusion of one’s selfishness and weakness.

I take this experience of sacrificing to be a sort of emotional performance, because I can think of no way to understand the difference between participation in the sacrifice and a lack thereof beyond the difference between genuinely emoting during the process of

\(^{21}\) *Analects* 3.12.

\(^{22}\) See *Analects* 11.12, in which Confucius responds to a question about serving spirits by asking how one could do that when one does not yet know how to serve people, and *Analects* 7.21, which explicitly says that Confucius did not talk about spirits. This also set them apart from the Mohists, who advocated a belief in spirits in part because they thought that such beliefs have practical benefits.

\(^{23}\) *Analects* 10.11.
sacrificing and merely going through the motions. I think this is what is also reflected in the statement that *junzi* are not mere vessels. Their role in the sacrifice is not one of an empty object, but of one who can actually go through the ritual with the proper emotional performance. This comports with the way in which we teach children to perform basic rituals like saying “please”, “thank you”, and “I’m sorry”. We do not want them to simply say “I’m sorry” as some sort of social necessity; we want them to *mean it*. One way that we teach them to mean it is by making them say it with the proper tone and facial expression that indicates the presence of an emotion of contrition. If my niece does not participate in the apology, it is like she has not apologized at all.

This example of children and apology rituals can help us understand how the *Analects* focus on Confucius’ composure and countenance when performing various ritual functions speaks to the function of the *li* in enacting certain emotions. His countenance would become battle-like when serving as the lord’s envoy, would express awe when walking passed the duke’s empty throne, and would take on a solemn appearance at the sudden sound of thunder, on being presented with a feast, and on encountering somebody in mourner’s attire. This focus on one’s countenance is not simply a matter of one’s face needing to be a certain way, as one’s face is an outward sign of one’s inward state, especially if one is being sincere and avoiding putting on airs and creating false appearances, as Confucius vehemently recommends. In these passages, then, we can see Confucius

---

24 *Analects* 2.12.
25 *Analects* 10.5.
26 *Analects* 10.4.
27 *Analects* 10.25.
28 This concern with being able to tell the difference between genuine earnestness and mere appearance can be seen in *Analects* 11.21 and 16.4. In addition to being an outward sign, I argue in Chapter 5 that there are
working to ensure that he was doing his utmost in using the available opportunities to practice feeling appropriate emotions at appropriate times.

The use of music in rituals also helps us to see how they are designed to enact a sort of emotional change in a person. Rituals and music are often mentioned together within the text. When they are, they are typically mentioned as though they share something:

The Master said: “What has a person who is not ren got to do with li? What has a person who is not ren got to do with music?”

Both ritual and music are activities that are connected to a persons having, developing, or exercising ren. While music is presented as distinct from ritual, I think that Confucius’ focus on music is a clear extension from the li proper to what I have called the broader ritual strategy in the Analects; enjoying music of a certain type in a certain way is taken to be a transformative practice in the Analects.

This “music as cultivation” idea comes through fairly clearly in passages that tell us that music helps create a “complete person” augmenting the developmental process of which the li are a part. Finding enjoyment in ritual and music is said to be advantageous, which is perhaps why Confucius warns us against seeing ritual as a matter of gems and silk and music as a matter of drums and bells. They are part of a similar transformational process, a process that we can see as important, I would argue, as a result of its emotional impact. Confucius is clearly emotionally impacted by music, as he once experienced music

---

biofeedback mechanisms that allow one’s performed facial expressions to impact one’s emotional state, and that this makes our ability to choose our facial expressions an effective intervention point for shaping our emotional responses both in a relevant moment and over a course of years.

29 Analects 11.1, 11.26, 13.3, 16.2, and 16.5 all have the two characters mentioned in the same phrase.
30 Analects 3.3.
31 Analects 11.12.
32 Analects 16.5.
33 Analects 17.11.
that was of such heights that he did not eat meat for three months.\textsuperscript{34} We are also told that during the three year mourning ritual, a \textit{junzi} would not enjoy the pleasant music that they may hear,\textsuperscript{35} and that on a day in which Confucius had wailed in grief, he would not sing at all.\textsuperscript{36} This tells us that the role of musical performances, to Confucius, is to have a certain emotional experience, an acceptable and necessary part of one’s pro-social development, but one that can be distracting at times of mourning, when, as I have argued, one ought to be practicing a different set of emotions.

Aside from the component of it that is relevant to music, the mourning ritual itself gives us perhaps our greatest insight into the emotion-shaping element of the \textit{li}. A mourning period of three years is long, especially compared to the natural and typical course of human grief after a loss.\textsuperscript{37} During that period, people are to wear special, drab mourning clothes, eat course grains, stay in a special mourning shack rather than normal lodgings, express their grief regularly, and, as I mentioned, take no joy in music. Why go through all of this? Well, one reason might be that it is what one needs to do to properly grieve. Confucius seems too taken with the notion of diligent effort to advocate such a long allotment of “personal time”, though. Another reason is that it is a duty of sorts. In criticizing Zaiwo for his comfort with a one-year mourning period, Confucius cites the three years of care that Zaiwo received from his parents when he was a helpless infant, before he could actively display filial piety in return.\textsuperscript{38} I think that this cannot be the entire story,

\begin{itemize}
\item \textsuperscript{34} \textit{Analects} 7.14.
\item \textsuperscript{35} \textit{Analects} 17.21.
\item \textsuperscript{36} \textit{Analects} 7.10.
\item \textsuperscript{37} See George A. Bonanno, \textit{The Other Side of Sadness} (New York: Basic Books, 2009) for an argument that humans in general are more resilient in the face of loss than even the popular “stages of grief” theory implies.
\item \textsuperscript{38} \textit{Analects} 17.21.
\end{itemize}
however, because the mourning ritual is a fairly central Confucian practice, and because the
typical age at which a person will end up mourning one’s parents is still young enough that
the person will be in a vital stage of their development (especially considering how the
Confucian school was geared toward moral education of adults). One does not cultivate
oneself so that one can ultimately fulfill this heavy commitment to one’s dead parents, and
Confucius did not mourn his parents after being able to give his heart and mind free reign
without overstepping the line. It is, then, more likely to be a significant part of the path
toward the type of moral competence at which the Confucians aim than simply an important
end in itself.

In light of these considerations, I think it makes sense to see the mourning period as
a period in which one is continuing to cultivate one’s person. It is a three year master class
in putting on the right countenance, in focusing and practicing feelings of familial love and
the motivation toward care and service that these feelings engender, in tamping down on
selfish urges toward comfort and pleasure that might make one think of personal gain before
they think of what is morally correct. One final piece of support for this reading comes
from the fact that the abilities that this mourning period can allow one to practice are shown
in the text to be of a high value to Confucius. He praises Yan Hui for his ability to enjoy
similarly uncomfortable (although, poverty-induced) circumstances without ill will,39 and
remarks that people who would feel ashamed to try to do so are not worth talking to.40

39 *Analects* 6.11. In 8.21 he gives similar praise to the sage king Yu.
40 *Analects* 4.9. If we see this comment about such being not worth talking to in light of other places in the
*Analects* in which Confucius encourages his students to seek out people who are *ren*, we can see this not as
merely a negatively-valenced judgment of such people, but as practical advice: such people will not help you
along this path of cultivation you are on.
This brings me to the third way in which Confucian rituals serve as a form of moral cultivation, namely by strengthening the kind of conscious control, i.e. executive functions, that can allow one to make reflective decisions to act in accordance with one’s moral beliefs. This building of one’s conscious control comes in part through the type of attentional and emotional training that I have already discussed. Paying attention to the details of ritual and to the emotional content of one’s face (and one’s xin (mind-and-heart)) requires a fairly high degree of explicit mindfulness and intent, and thoughtless performances will not instill the right sorts of habits.41

This strengthening of our executive capacity is, I believe, what is largely behind Confucius’ focus on zhong. Etymologically, zhong shows a centered mind-heart, which we can take as indicating diligence, loyalty, or focus. The meaning posits suggested by the etymological clues comport with how “zhong” is used throughout the text, and its use throughout the text supports the idea that part of the Confucian method of self-cultivation includes building one’s capacity to focus one’s efforts. We are told that zhong, along with culture, conduct, and making good on one’s word, were the four things that Confucius taught.42 Being conceptually separated from culture and conduct in that way lends support to the idea that zhong is not merely a part of how one should conduct oneself, or one of a list of virtues to be had, or a part of the Zhou culture that Confucius is revitalizing. What it is instead, I think, is an approach, both to one’s moral education, and to how one conducts one’s life in general. It lies outside proper conduct in the way that focused efforts lie outside

41 As any of my percussion students know, there is a huge difference between practicing so as to instill good habits and simply hacking through parts.

42 Analects 7.25.
of the rules of a sport or the principles behind playing a musical instrument. It is a sort of all-purpose setting for one’s pursuit of one’s goal.

I think this reading can be borne out by analyzing a passage that also speaks to the centrality of zhong in Confucius’ method:

The Master said, “Zeng, my friend, my dao is bound together with one continuous strand.”
Master Zeng replied, “Indeed”.
When the Master had left, the disciples asked, “What was he referring to?”
Master Zeng said, “The dao of the Master is zhong and shu, nothing more.

I discussed in footnote 154 of Chapter 3 how shu is sometimes taken as a sort of Confucian golden rule; some take the passage here to indicate that zhong somehow adds content to this golden rule, and that this rule then comprises the way of Confucianism. Feng Yu-lan, for instance, sees zhong as a positive element of that one golden rule. Philip J. Ivanhoe has argued persuasively that such a reading lacks support and rests heavily on reading the content of later texts back into the Analects. His view, however, that zhong is a strict form of rule following in service to leaders, I find too strong of a reading for the available evidence. He cites a passage that implores one to be zhong in service of one’s leaders and

---

43 Analects 4.15. There has been some argument that this passage is apocryphal. See, for instance, Bryan Van Norden, "Unweaving the 'One Thread' of Analects 4:15," in Confucius and the Analects: New Essays, ed. Bryan Van Norden, 216-236 (New York: Oxford University Press, 2002). I find the arguments convincing enough to avoid seeing this passage as spelling out the key to early Confucianism, and possibly to see this passage as a bit of pro-Zeng advocacy on the part of later students under his influence, but I still consider this passage to be consistent enough with what we find elsewhere in the text to not rule it out as a piece of textual evidence completely.
says that in this passage, “we clearly see the hierarchical nature of the concept zhong”\textsuperscript{47}. I do not agree that the hierarchical intent is clear here, as this statement is merely a response to a question about how rulers should employ ministers and ministers should serve their lord, and one can say that others should be zhong in serving their leaders without implying that zhong is an inherently leader-oriented way of conducting oneself.\textsuperscript{48} This hierarchical reading of zhong also leads a problematic reading of 13.19, which says to be zhong in how we relate to others after saying to be respectful in handling one’s official affairs. Not only does the passage seem to allow that one can be zhong toward all people, and not just one’s peers and superiors, but it also ends with the statement that one could not do without such an attitude even among the Yi and Di barbarians, who we can safely assume do not have an established hierarchical li system within which to be zhong.\textsuperscript{49}

So how, then, does zhong actually fit into the one thread? I think, first off, that if we take the passage seriously in saying that there is one thread, it would be odd to say that it had “two strands”, like zhong and shu conceived as separate tasks, methods, etc. I think we would do better to parse the phrase “夫子之道忠恕而已矣” (The masters dao is zhong shu and that’s it) so that instead of listing zhong and shu as two separate methods that come together

\textsuperscript{47} Ibid., p. 25.
\textsuperscript{48} There are, of course, other passages that use “zhong” to describe a way that one should be in serving rulers or in being a friend, such as 2.20, 5.19, and 12.23, but I think that these passages only establish that one can be zhong in these situations, not that these situations are the only ones in which one can be zhong, or even that these are paradigm cases of zhong. Cf. Sin Yee Chan’s parallel denial that shu is hierarchical in nature, in Sin Yee Chan, "Can Shu Be the One Word That Serves as the Guiding Principle of Caring Actions?," Philosophy East and West 50, no. 4 (2000): 507-524, endnote 11. Cf. also Paul R. Goldin, "When Zhong Does Not Mean Loyalty," Dao 7 (2008): 165-174.
\textsuperscript{49} Analects 13.19. I think a similar argument could be made with reference to 12.14, in which Confucius responds to a question about governance by saying to “reside in it without tiring, and carry it out with zhong”. Governance here (zhong) does not seem to be a matter of serving one’s superiors or peers, because it is so frequently discussed in the Analects as involving leadership activities.
under some unstated central theory, we can see “zhong 忠” as modifying “shu 德”. The one thread, on such a parsing, would be to “diligently shu.” If we consider this the kind of advice aimed at novices, which is most likely who would have been looking to the relatively young Zeng for his take on Confucius’ claim, we can see this one thread as a key piece of advice for a particular stage in one’s development. On such a reading, zhong remains a type of loyalty to an end, so there is no mystery of how to square this passage with the other usages of “zhong” in the Analects. It is not something to do, but a manner in which one can do something, a manner of doing things that Confucius is trying to instill in his students because such diligence will always be helpful in making one’s endeavors, including one’s moral endeavors, successful.

In addition to the passages that mention zhong, we can also see a general focus in the text on learning how to maximize one’s efforts, how to be in control of what and how much one is doing to ensure that one is doing things correctly, comes across clearly as one of the Confucians’ goals in the text. His criticism of Zaiwo for sleeping during the daytime indicates a value on diligent effort, as does his claim that he can do nothing for those who do not ask themselves “What to do?, What to do?” When one student decides that Confucius’ way is simply beyond his strength, Confucius says that those lacking strength collapse further along the path; the student is drawing his own line, succumbing to his inclination toward laziness rather than putting in the required conscious effort to change. This focus on one’s effort can also be seen in how rituals often forced a person to do things

---

50 Analects 5.10.
51 Analects, 15.16.
52 Analects 6.12.
the hard way rather than the easy way. Fishing with a line rather than a net and not shooting at roosting birds makes these activities more difficult, which would make little sense if the ends of these activities were the only goals. The goal was most likely not to preserve animal welfare, so it makes sense to see the goal as making the activities exercises of effort, self-control, and skill.

The type of mindfulness and making sure that one is putting in the effort one must in order to do things correctly that I have been discussing is also on display in *Analects* 5.20:

Ji Wenzi only took action after thinking about it three times. On hearing of this, the master said, “Twice would suffice.”

Confucius is here encouraging an overly cautious student to be less so, but the method that he leaves in place is still one in which one must be thoughtful in everything one does. This habit of thinking before one acts, I would argue, is intended to keep people from transgressing by giving them an opportunity to exert some effortful control over their behavior. This, I would further argue, works in concert with performances of the *li*, in which one must be mindful of a great deal, and with ritual strategy methods like asking oneself three questions daily about how one has conducted oneself, to keep one mindful of one’s conduct, and to strengthen one’s executive faculties to help one stay in control of one’s behavior.

The fourth effect of rituals, that of shaping our environments in ways that aid cultivation, has both what could be called a physical dimension and what could be called a

---

53 *Analects* 7.27.
54 Confucius, famously, neglects the wellbeing of horses after a fire in *Analects* 10.17, and cares far more about ritual than the plight of a sacrificial sheep in 3.17.
55 *Analects* 5.20.
social dimension, although the two are clearly coextensive in various ways. The physical
dimension includes the prescriptions for how one should decorate one’s house, how one
should dress, what sort of architecture places important rituals take place should have, how
the mats are to be set up at meal time, etc. We can see in the Analects that great care is taken
to make sure that one’s surroundings are organized in a certain way. If, as the Analects says,
the greatest value of the li is in creating harmony, and many of the li tell us ways that we
need to structure our environments, I think it makes sense to see an understanding in the
Analects that one’s environment has an effect on whether harmony occurs, and that the li
help us produce environments that serve to facilitate harmonious behavior.

As I mentioned, though, there are parts of this physical dimension that involve other
people, and so also contribute to the social dimension. Other people’s facial expressions,
posture, movements, proximity, etc. all contribute to the basic perceptible environments that
we find ourselves in, and these sorts of things are prescribed in various li. A day of moving
through a world in which people are following the li, controlling their expressions in certain
ways, deferring to others, giving solemn offerings to ancestors, etc., is simply different than a
day of moving through a world in which these things are not taking place, and the Confucian
focus on the li seems to imply that this difference is important.

---

56 Analects 1.12.
57 As I will argue in the upcoming chapter on the empirical support for a Confucian method of moral
cultivation, we now have good reason to think that such environmental differences are of fairly great
importance in determining the types of behaviors that people perform. While it would of course be
implausible to think that Confucians were privy to the facts of situationist psychology, or that a defense of that
line of thought as we currently understand it exists within the Analects, the notion that structuring an
environment in one way rather than another is an important element in facilitating certain behaviors rather than
others does seem to operate in the text, is something that is robust enough to be plausibly apprehensible to
some degree by thoughtful and perceptive humans without modern science, and comports with a number of
There is something more to this social dimension, however, than merely giving us a certain environment to move through. The structuring of the social environment impacts how we interpret the events that are going on around us. It affects not only what a particular experience is (i.e., a person bowing in front of us), but also what we experience it as (i.e., a friend being deferential). This takes place in a number of ways, and it often does so through the prescriptions for how people in one role ought to behave toward people in another role. In order to understand this aspect of how rituals function, however, we need to look at the way in which social roles are woven into the fabric of the *li*, which I shall do in the next section.

4.3 Roles and “Rectifying Names”

Part of the way in which rituals serve to structure our environments is by determining what people ought to do when interacting with others. This process is aided by many of the rituals having built within their structure a separation of people into different roles. To exist in a role is to exist in a certain place in one’s social environment; roles thus help structure the social environments in which we find ourselves by giving us a way of framing what those environments are.

We can see how roles function in the text by first looking at how rituals tend to prescribe behavior for people based on their roles. The *Analects* show Confucius acting in general characterizations of Chinese thought, and Eastern thought generally, as more holistically and less individualistically minded than Western philosophy.
specific ways toward his lord\textsuperscript{58}, the duke,\textsuperscript{59}, the blind\textsuperscript{60}, higher and lower officials\textsuperscript{61}, and his friends\textsuperscript{62}. Passages discussing the death of Yan Hui offer perhaps the most vivid picture of how rituals countenance roles. Despite being Confucius' favorite student, Confucius told Hui’s father that he would not sell his carriage to ensure that Yan Hui had an outer shell for his coffin. This was both because Hui’s father should treat his son as a son (who, in burial, gets a coffin but no outer shell, just as Confucius’ dead son had), and because Confucius’ role as a retired official made it improper for him to travel on foot. When Hui was buried, Confucius’ disciples threw him a lavish burial, despite Confucius’ remonstrations that such a burial would not be fitting. In expressing his disappointment, he claims that Hui treated him as a father, but Confucius was not able to bury him the way a son ought to be buried according to the \textit{li}.\textsuperscript{63}

Attributing roles to people serves a function of simplifying the social world we inhabit. Interacting with people qua the individuals that they are is taxing, and makes competitions for resources, affection, and standing within a family a war of all against all.

\begin{footnotesize}
\textsuperscript{58} \textit{Analects} 10.18-10.20
\textsuperscript{59} \textit{Analects} 10.2.
\textsuperscript{60} \textit{Analects} 9.10
\textsuperscript{61} \textit{Analects} 10.2.
\textsuperscript{62} \textit{Analects} 10.22, 10.23, 10.15.
\textsuperscript{63} I discuss this further on in more detail, but, to plant the seed of my later argument, I'll just say here that, considering Confucius and Yan Hui were not father and son, it seems interesting that Confucius would describe their relationship in those terms, in a way that was clearly meant to be action-guiding. I think this lends support to the idea that roles themselves (or rather, the act of role attribution) are a part of the Confucian approach to moral cultivation. The view that \textit{ren} has a familial root would suggest that there is something in fulfilling familial roles that effects a person in a way that makes them more pro-social, and a natural strategy from here, to make it more likely that people will perform their duties, is to try to extend what takes place within family roles to what takes place within the broader community. I believe that the Confucian strategy for doing that was to use familial role attributions in the wider society. The person who serves the ruler properly is the one treats the job like the ruler’s son. The ruler is the parent; somebody has to cut the turkey, lead the sing-along, and hand out the gifts. When one takes up an official position, one attributes to oneself the role of the \textit{junzi}, on the assumption that this kind of attribution makes one better able to fulfill the duties of the role.
\end{footnotesize}
Children who are given unequal treatment by parents learn to compete with one another, and develop feelings of enmity to the detriment of feelings of affection. Attributing the role “parent” to oneself and “son” or “daughter” to all of one’s children serves to simplify matters like how expensive of a gift a child gets on his or her birthday, how much desert everyone gets, etc. Standardization makes competition fruitless, and so the seeds of enmity, competition, and selfishness do not take root. The elli for burying a son has this sort of equalizing effect that helps to make a family the kind of place where pro-social feelings and habits can take root. The children serve the parents according to the elli, and practice their pro-social feelings and habits. The parents serve their children according to the elli, and thus provide the space necessary for that practice of pro-social feelings and habits. Somebody needs to be in charge of applying the moral technology, both in families and in the society. People “downstream” in the process need people “upstream” in the process to apply it, which will enable a sort of development that will help keep them from going astray due to habits not conducive to correct behavior. This also helps to make people more capable of fulfilling the upstream role when they get there. In a functioning family, the parents are further upstream in this process than their children; in a functioning society, the rulers are further upstream in this process than the people.

There is more to the notion of roles in the text, though, aside from the structure that they help provide in the elli. Role attributions themselves appear to be a part of the Confucian vision of what will lead to effective moral development. We can see this, I believe, from the passages that are typically cited to support the idea that the “rectification of names” is a central feature of Confucianism. While I agree that rectifying names plays an important role,
I consider the naming that Confucius advocates to be a type of role attribution. On my view, this role attribution, and the embodiment of these roles in the *li*, makes up an important part of the Confucian approach to cultivating compliance to one's moral standards.

The passage that mentions the rectification of names has Confucius responding to the question of what he would do if handed over the administration of a state. He responds that he would rectify names, and when challenged, he explains:

“When names are not used properly, language will not be used effectively; when language is not used effectively, matters will not be taken care of; when matters are not taken care of, *li* and music will not flourish; when *li* and music do not flourish, the application of laws and punishments will not be on the mark; when the application of laws and punishments is not on the mark, the people will not know what to do with themselves.”

The rectification of names, in this passage, sets the stage for everything else that comes after to happen correctly. It is clearly important, and its importance raises some interesting questions. One such question is what names need to be rectified. Considering the way in which philosophy of language developed in pre-Qin philosophy, I highly doubt that Confucius was arguing for a form of scientific realism where everything in the world must be given its proper name. The names that need to be rectified, I think, are titles, or, in other words, names that can be used to attribute roles.

I think this for a couple of reasons. One reason is the existence of a story in the *Zuo Commentary to the Spring and Autumn Annals* of, as put in a footnote to Ames and Rosemont’s translation to the *Analects*, “Wey rewarding a commander who had come to their aid, giving

---

64 *Analects* 13.3
him the use of musical instruments and the right to appear at court using emblems of a prince\textsuperscript{65}, after which Confucius criticizes this infelicitous attribution of an official title as a mistake that could cause everything to fall apart. The concern here seems to be with roles and taking their attribution seriously. The other reason is that \textit{Analects} 12.11, a passage that appears to compliment the passage that explicitly mentions the rectification of names, specifically mentions names for social roles, \textit{“jun jun, chen chen, fu fu, zi zi”} (unparsed, this amounts to “ruler ruler, minister minister, father father, son son”). This passage is thought to be about the rectification of names because it clearly involves concern for a fit between names and what those names signify. The direction of the fit is not made quite clear by the passage, as it could be parsed either as saying that leaders need to lead, ministers need to minister, etc., (so people need to fit their titles) or as saying to “leader” leaders, “minister” ministers, etc. (so that titles are used to order people). This latter parsing would imply that governing effectively is a matter of rectifying names, of being sure to apply titles in a consistent, appropriate way, and this would support the idea that Confucian name-rectifying is specifically concerned with rectifying role attributions.

If these role attributions are themselves important to the Confucian approach, then I believe that we can see the role-ladenness of the \textit{li} in a different light, in which the \textit{li}, in addition to the other goals that I have mentioned, serve as a vehicle for making and reinforcing role attributions. One is not simply dubbed a father once and then expected to act the way that a father should. Rather, the attribution of the role is reinforced in rituals that allow people to conspicuously experience their roles, and, presumably, to be continually

\textsuperscript{65} Ames and Rosemont, \textit{The Analects of Confucius}, p. 253.
reminded of the duties and characteristics of people who fulfill their roles in ways that accord with what is considered to be morally acceptable (yi). If rituals are largely ritualized role-performances, then there is further support for the idea that role attributions are a key component of the Confucian moral implementation program.

The hypothesis that, in the Analects, role attributions are seen as an effective tool for building moral compliance might yield a plausible understanding of the importance of rectifying names and highlight one of the mechanisms at play in Confucian rituals, but we should be able to say more about how this notion works in the rest of the text if we are to see that interpretive hypothesis as well-supported. Fortunately, I think that the role attribution hypothesis allows for an interestingly unified account of why the Analects focus simultaneously on family, moral development, the goal of becoming a junzi, rectifying names, and the career of public service.

One piece of this account is the idea that we first learn to be pro-social by doing our utmost in our family roles. This pro-sociality, or ren, is something that, if we were to carry it into the rest of our interactions, would help enable ethically correct behavior. Having the role of “son” or “daughter” attributed to one when one is young and being provided with the appropriately ritualized family space to practice this role helps build certain pro-social feelings and habits. Over time, being reminded that one is a son or daughter and performing rituals that serve as vehicles for this role attribution activates these feelings and habits. Activation of these feelings and habits motivates behavior that falls within what was considered within the bounds of yi, i.e., what does not overstep the line.
With insights into the benefits of the traditional rituals from both study and practice, Confucius, training those who will serve the leaders in official positions, developed the idea of the junzi as a role that people in official service should take up. This role, literally, is the role of the “ruler’s son”. One acts like the son of the ruler, because, having built certain habits around the role of being a son, using this role attribution helps activate the pro-social feelings and habits that one has developed, facilitating the kind of behavior one should exhibit, inhibiting urges to go astray. This strategy was also at work in Confucius’ school, as evidenced by story of Yan Hui’s funeral and the way in which it shows that Confucius and his students modeling their relationship on a father-son relationship.

In order for this role attribution to have the appropriate effect, though, roles and role terms must be preserved and protected. If the lines between the roles get blurred by people decorating their houses like the rulers, people wearing inappropriate title insignia, and so on, then the effects, Confucius fears, will be diminished. This is why the rectification of names is important. When words get used in a variety of contexts, their ability to consistently invoke what they invoke dissipates. A century ago, nobody would have thought to call a cup of coffee “awesome”. As it has been used more and more frequently, the power of its attribution has diminished. Words are like that, unless those who hold power over cultural trends help to preserve their force.

This is why governance is a matter of “son”-ing sons, “father”-ing fathers, “minister”-ing ministers, and “ruler”-ing rulers, and why, I believe, Confucius mentions these four roles specifically. They may just be the first handy examples, but I think a deeper reading than that, casting the choice of these roles as intentional, is clearly plausible. These
are the key role attributions that support what I see as the family-to-public strategy of Confucian moral cultivation. We need to make sure that the fu-zi relationship holds up, because it is the basis and model for successfully engaging in the jun-chen relationship. We train ourselves in our families such that these roles motivate proper behavior, and then we co-opt this effect in our public lives, by taking up the role of the son of the ruler, which, just as the parent-child relationship gives us a role as siblings, makes us a brother or sister to the people (which also helps make sense of Zixia’s claim that everyone in the world is a brother to the junzi66). The right way to be a minister is to take up the role of the ruler’s son rather than a mere functionary (a vessel), and the right way to be a ruler is to simply be a father to the people. In order for this to work, though, we need to make sure that those most conspicuously attributing and enacting those roles are doing so consistently.

One possible objection to this view is that the focus on rectifying names appears to be at odds with the co-opting of family roles. Why would rectifying names come as a pre-requisite to applying names in the new, co-opting way? Does the co-opting of role terms not defeat the purpose of rectifying them? My answer to this objection is that there is a difference between lapsing in applying terms like “father” and “son” by attributing them to people who, say, abuse one another, and applying it to somebody that serves the ruler as one would serve a father. One blurs the application conditions for the terms “father” and “son” such that they fail to invoke the proper feelings and habits. The other applies it strictly to a

66 While, as I mentioned in chapter 3, I have my issues with Roger Ames’ interpretation as a whole, I do think that there is an interesting connection between this strategy of co-opting family roles as a way of bringing sympathy into public life is reflected in, for instance, in using “big family” (dajia) to refer to “everyone”.

166
person performing a public role in a manner modeled on the roles “father” and “son”.

There may be some surface tension, but it does not create a problem for the strategy.

There is also tension, though, between this strategy, which appears to instill a form of impartial care, a “loving others as one loves one’s family”, which clearly runs against the Confucian history of advocating for partial concern. There is evidence in the Analects that even Confucius advocated partial concern; when one’s father steals a sheep, one should cover for him rather than doing what is in the public’s best interest. If role attributions are intended by the Confucians to work as I have said, have they not set up a method of cultivating moral compliance that would fail to instill habits conducive to fulfilling whatever portion of their moral commitments are partialistic? And does that not count against the coherence and thus the charity of my interpretation?

I do not think so, because I do not think that the inevitable end of the strategy that I have laid out is ever going to be equal concern. I think it makes sense to think that, given the Confucian private-to-public strategy, there will always be a lag between the strength of what allows one to fulfill one’s partial concerns and what allows one to fulfill one’s impartial concerns. To whatever degree my taking up the role of a “brother” to the people serves to motivate acceptable public behavior, it will be because at least some of the habits that I have built in being a brother to my siblings (an easier job) are activated by taking up the role of “brother” to the people (a more difficult job). There is no reason to think that role attributions in this extended context will be as effective in activating these feelings and habits as those attributions in their normal context, much less overpower those normal-context

---

67 Analects 13.18.
attributions. Using this moral implementation method, one’s ability to care for one’s family continually improves, and on its coattails rides one’s ability to serve the public.

4.4 Conclusion

Having argued in Chapter 3 that interpretations of the Analects that attempt to see it as presenting and defending a particular, systematized normative ethic all run into various problems, either by failing to account for parts of the text in a consistent manner, or by looking at the Analects through later-Confucian lenses that ignore diachronic philosophical development, I presented, in this chapter, my argument that the project of the Analects is to present a method of implementing a morality, a way of cultivating oneself into the kind of person that does not succumb to transgressing one’s moral lines. I argued that this moral implementation method involves the use of rituals and role attributions to build a greater tendency toward and capacity for keeping one’s conduct within acceptable bounds. This interpretation, I have tried to make the case, makes better sense of the totality of what we find in the text of the Analects, and the place that the Analects fits into the history of the pre-Qin philosophical conversation, than do the various normative ethics interpretations. It also, as I will argue in the next chapter, offers us a method of moral cultivation that we have good reason to believe can yield results for modern people who wish to better accord with the moral standards that they accept.
Chapter 5: On the Effectiveness of the Analects’ Method of Moral Cultivation

5.1 Introduction

In Chapter One, I made the case that normative ethics needs to proceed in light of the best available picture we have of our psychology, and that this best available picture should include an account of our best available methods of moral cultivation, i.e. the processes by which we can shape and maintain our psychologies so as to come to more fully and consistently accord with our moral norms. In Chapter Two, I presented an empirically-supported picture of some of the primary impediments standing in the way of full and consistent accordance with our moral norms, a picture that helps to tell us why we need methods of moral cultivation in order to comply with our moral norms and what we need them to do for us to help secure that compliance. In Chapters Three and Four, I made the case that on the best reading of the Analects of Confucius the text provides us with one such method of moral cultivation, and I explained how that method serves to move us toward greater compliance. In this final chapter, I will make the case that we have good reason to believe that the method of moral cultivation that we find in the Analects is an effective and distinctive method, and that the Analects can thereby make a valuable contribution both to discussions of ethics as a practical endeavor and to a suitably naturalistic approach to normative ethical philosophy.

We do not, unfortunately, have what I would consider to be the best sort of empirical evidence that I would need in order to argue that the Confucian method is
effective, i.e., studies showing that people pursuing a ritual-based strategy intended to shape our emotional and cognitive capacities in certain ways and load our self-concepts with behaviorally-relevant and frequently-referenced role terms end up behaving in a more consistently norm-compliant manner;¹ those studies have simply not been performed yet. In light of this lack of direct evidence, what I shall do here is to present a philosophical argument, one that draws on currently available empirical evidence regarding human psychology and the behaviorally relevant ways that we can shape and affect it, that attempts to show that we have good reason to think that the practices and strategies that make up the Confucian method can produce the intended effects. If my argument is successful, we should then be justified not only in pursuing further, more specific research into the effectiveness of these strategies, but also in pursuing these strategies in our own lives and advocating them more broadly, at least until new evidence alters the justificatory landscape.

Then, in order to show the further point that the *Analects*, beyond giving us a method that we should believe can work, makes a unique contribution to the current state of naturalistic ethical discourse, I will further argue that the method gives us something valuable that we do not already have, i.e. that the method is likely to be either more effective, or more desirable in some other ways, than other methods of moral cultivation that we can already find as a part of philosophical discussions of cultivating compliance to our moral norms.

The rest of the chapter is divided into two sections. In section 5.2, I will present the empirical support for the Confucian method, showing that the practices that make up the

¹ Recall, as I noted in Footnote 3 of Chapter Two, that for the purposes of this project, I am not taking any particular stand on what we should take to be the best normative theory. Obviously, in arguing that a particular method counts as an effective method of moral cultivation, we need to assume some dividing line between morally correct and morally incorrect behaviors. Here, I rely on a sort of minimal moral consensus.
method are the types of practices that we have good reason to think will have the psychological effects that I claimed in Chapter 4 that they are intended to have, and I will make the case that, in having these effects, the method should help us to overcome the impediments to consistent moral behavior that I outlined in Chapter 2. In section 5.3, I will lay out a few attempts in the Western literature to bridge the gap between where we start psychologically and where we need to be in order to behave in a consistently moral manner. I will conclude by arguing that the cultivation method in the Analects makes a unique contribution, by showing that the Confucian method performs better than these other methods according to the relevant assessment criteria, which includes not only the effectiveness of the method, but also other criteria including the practicability and attractiveness of the method.

5.2 Support for the Analects’ Method

In this section, I will attempt to show that we have reason to believe a number of individual claims that, collectively, should lead us to believe that the Confucian approach should be effective in cultivating moral behavior. I contend that if I can establish that we have good reason to believe these claims, then we have good reason to think that the Confucian method found in the Analects, as I have laid it out, is an effective method of moral cultivation. Those claims are as follows:

(1). Our attentional capacities and tendencies are trainable through ritual activities.

(2). Training our attentional capacities and tendencies can improve our ability to comply with our moral norms.
(3). Our executive functions can be strengthened through ritual activities.

(4). Training our executive functions can improve our ability to comply with our moral norms.

(5). Our capacity to have insight into our emotions and moods and the emotions and moods of others is trainable through ritual activities.

(6). Increasing our capacity to have insight into our emotions and moods and the emotions and moods of others can improve our ability to comply with our moral norms.

(7). We can alter the content and/or strength of our emotional responses and moods through ritual activities.

(8). Altering the content and/or strength of our emotional responses and moods can improve our ability to comply with our moral norms.

(9). Role attributions affect our ability to comply with our moral norms.

(10). Rituals that reinforce how we understand our roles serve to make role attributions more effective.

(11). The sort of structure that rituals can give to our environments can improve our ability to comply with our moral norms.

I will proceed in as linear a manner as these claims allow. This is because the support for some of these claims also serves to support certain other claims, since, as I will discuss, the neural processes and structures that underlie our attentional, executive, and emotional-regulatory abilities appear largely to overlap; as a result, training our attentional
capacities can improve our general executive capacities, and changing our attentional and executive capacities can have effects that influence our moods and emotional reactions, and vice versa.

A good deal of the evidence that can be cited in favor of claim (1) through claim (8) can be found in the vast literature in recent psychology and neuroscience discussing the various benefits of attentional training tasks that collectively go by the name “mindfulness”. Mindfulness “incorporates elements of both attention-regulation and an open, accepting orientation to experience”\(^2\); we can see that openness as an attention to our immediate emotional and cognitive responses to our experience along with a plan to not act on them when they occur. Mindfulness training is used to develop mindfulness skills, and involves engaging in various practices of attending to one’s experience in a certain way, whether that means a spending time performing activities with a detached awareness of one’s thoughts and emotional responses, a focus on one particular stimulus like one’s breath or one’s body, or a focus on one particular thought or emotion. Mindfulness practices are influenced by Buddhist meditation practices, but have been adapted recently to a secular Western paradigm\(^3\) as a part of therapy programs like dialectical behavioral therapy\(^4\), wherein “‘Meditation’ may be understood simply as attentional training.”\(^5\) These mindfulness-based therapies are currently used for the treatment of a variety of illnesses and ailments, such as

---


borderline personality disorder, depression, chronic pain, anxiety, eating disorders, and are also taught as a technique for reducing stress.

My argument that Confucian rituals are effective rests on this literature not simply because the literature shows that the kinds of effects that I claim Confucian rituals produce are possible, but because, as my discussion of the Analects in Chapter 4 made clear, we can see that something incredibly similar to this kind of mindfulness is an important element of the Confucian ritual training program. Among psychologists studying mindfulness practices, “It is important to note that mindfulness refers to a particular quality of attentional focus, mindful awareness, rather than to any particular practice or technique”12, and I think that the Confucian practices we find in the Analects share that particular quality of attentional focus. Just like the various mindfulness practices drawn from Buddhism or developed in therapeutic settings, training via Confucian rituals, i.e. being present in a sacrifice, greeting people in the right way, taking on a solemn countenance at the right time, etc. often means

6 Linehan, *Cognitive-behavioral Treatment.*
directing our attention to our emotional and bodily states and responses or being in a similar state of presence and awareness.

5.2.1 Attention – Claims (1) and (2)

A belief in claim (1), that our attentional capacities and tendencies are trainable through ritual activities, gains at least some initial support from our common experiences of learning to play musical instruments, to play sports, to play games like chess, to drive a car, to cook, or any of a number of other skills. To be successful in any of these sorts of endeavors, just as in the case of social rituals, we need to learn to attend to certain elements of our experience and ignore the rest. Part of the training that I give when I teach percussion, for instance, is intended to teach my students how to direct their attention to the right aspects of their experience so as to be aware of their technique, the sounds and rhythms that they are producing, whether they are with the metronome, and whether what they are each playing is lining up with everyone else. Over the course of a summer, they learn what to pay attention to and develop the ability to do so more correctly, more consistently, and in ways that seem to them to occur more automatically. This greater automaticity of certain attentional tendencies is something we experience and rely on when we, for instance, drive while talking on the phone or while eating a burrito; we are able to focus on these other activities without crashing into things because over time we no longer need as much directed focus to be able to attend to the driving-relevant aspects of our experience, like street signs, other cars, when we need to signal, and so on. As I discussed in Chapter 4, training by way of Confucian rituals, like learning these other skills, includes a
focus on cultivating the same sort of attentional capacities and tendencies, leading practitioners to pay attention to, for instance, their posture, their breath, their facial expressions, and the emotional state that they are exuding through the performance. So, it seems like there is at least some foundation in our common experience upon which to rest these claims.

Because a part of the way in which our attentional tendencies and capacities are shaped involves certain attentional processes becoming more automatic, we will be able to get a firmer handle on that shaping process if we have a clearer understanding of what it means for an attentional tendency to become automatic. In what ways are attentional skills becoming more automatic as we perform these skills, and how does this affect the ease with which we can come to perform tasks that require attentional skills? To experience some task as automatic typically involves experiencing an absence of what goes into deliberate, controlled action, namely awareness, agency/intention, effort, and control. Psychologist John Bargh has argued that rather than think of automatic processes as all similarly exhibiting an absence of these four qualities (or, if you like, exhibiting the qualities of unawareness, unintentionality, ease/efficiency, and being uncontrolled), we should consider these to be “separate, distinct qualities that can be true of any cognitive process and can combine in a componential fashion to place processes along an automatic-controlled continuum.”

---

There are a couple of conceptually-different processes that we can see coming to exhibit these qualities in the performance of tasks that require a particular pattern of directed attention. One is the process of attending, positively, to the right aspects of one’s experience; we are presented with a pattern of stimulus, and our attention gets drawn to the right things. In learning to play music for instance, there are a number of aspects of one’s performance that one must get right, and one needs to learn to pay attention to all of it, from tempo, dynamic, rhythmic accuracy, and intonation, to sound quality, expressiveness, and phrasing. As this positive attention becomes more automatic, one becomes better able to listen to a musical performance and notice these aspects without effort, without forming an intention, without being actively aware that one is processing that particular information, and without control.\textsuperscript{14} Another closely-related process that becomes automatic is keeping one’s attention from being drawn by irrelevant, distracting elements of one’s environment, so as to continue to attend to the right sorts of things. A person who has mastered the skill of driving a car in a busy city, for instance, has learned how to tune out a vast array of potentially attention-drawing stimuli so as to still attend to the right signals, the flow of traffic, and so on.

Being able to pay attention to certain things automatically, and to avoid paying attention to other things automatically, can also make it easier to pay attention to more and more of what is relevant in the performance of a skill. As Bargh explains, “the automatization of routine thought processes frees one’s attentional resources for non-routine matters and enables a reduction of the massive amount of stimulation and

\textsuperscript{14} This lack of control over automatic attention processes, unfortunately, has ruined many of my friends’ ability to comfortably listen to rock bands who play without consistent tempo.
information bombarding one at any given moment into a more manageable subset.¹⁵ When we learn, for instance, to automatically notice certain patterns on a chess board, or on the other hand, to ignore parts of the board that are no longer where the action is, it allows us to more effortfully attend to deciphering an opponent’s strategy, or when we come to automatically pay attention to maintaining tempo and dynamic balance while playing music, or, on the other hand, to ignore certain elements of the ensemble, it allows us to pay more attention to aspects of the performance like expressiveness. In this way, making certain attentional tendencies automatic is one way to extend our capacity to attend to all of the things that are relevant to the performance of a skill.¹⁶

Looking into the recent literature in psychology and neuroscience, we can see that there is plenty of evidence to support the idea that certain training methods can improve our attentional capacities and instill in us various attentional tendencies. In one study, the experimental condition involved a three month training program in focused attention meditation, and participants’ attentional capacities were tested in an activity in which they “were instructed to attend to tone pips presented in one ear (e.g., the left ear) and to press a button each time they detected an intermittent deviant tone among the frequent (“standard”) tones in that ear [and] to ignore concurrent tone pips in the opposite ear.”¹⁷ Those who went through the training program showed “reduced variability in attentional processing of target

---

¹⁶ I discuss in section 5.3.1 some further ways in which automaticity helps to aid in the performance of certain skills, i.e. by automating certain sub-skills so that we can competently perform them without needing to attend to them, freeing our attentional resources for more advanced aspects of the skill.
tones.”\(^{18}\) This reduced variability was shown both behaviorally, by a reduced variability in reaction times and by “enhanced theta-band phase consistency of oscillatory neural responses over anterior brain areas”\(^{19}\), which indicates that the neural process underlying the attentional task was itself becoming more consistent. The behavioral and neural differences were also positively correlated with one another, as “those individuals who showed the greatest increase in neural response consistency showed the largest decrease in behavioral response variability.”\(^{20}\) This suggests that training to improve the consistency of our attention will have results, and that it will have those results because such training can affect how our brains actually function.

Additional support for the idea that attention is something that we can train through mindful practices like Confucian rituals can be found in studies on the neuroscience of attentional practices. A recent study that compared practitioners of mindfulness meditation with the general population “confirmed greater gray matter concentration for meditators in the right anterior insula, which is involved in interoceptive awareness.”\(^{21}\) A greater concentration of grey matter in a particular area means that more neurons are available and dedicated to performing the function of that area of the brain, which, given recent studies that have begun to correlate this kind of structural difference with functional differences\(^{22}\),

\(^{18}\) Ibid., p. 13418.
\(^{19}\) Ibid., p. 13418.
\(^{20}\) Ibid.
strongly suggests both, generally, that alteration of concentrations of grey matter is one of the mechanisms underlying our neuroplasticity, and, specifically, that the practices of these meditators improve their attentional capacities in part by inducing functionally-relevant structural difference over time. It was also found that the “mean value of gray matter concentration in the left inferior temporal gyrus was predictable by the amount of meditation training, corroborating the assumption of a causal impact of meditation training on gray matter concentration in this region.”23 Engaging in the practice of directing one’s attention toward one’s mental, emotional, and bodily states appears to strengthen the neural underpinnings of those processes.

Even shorter term courses of training have yielded significant results. In one study, novice meditators participated in a ten-day mindfulness mediation retreat in addition to completing various cognitive and affective assessments. The results of the study “indicated that those completing the mindfulness training demonstrated significant improvements in self-reported mindfulness, depressive symptoms, rumination, and performance measures of working memory and sustained attention, relative to a comparison group who did not undergo any meditation training.”24 In another study, participants engaged in Integrated Body-Mind Training (IBMT), a method that brings together “several key components, such as body relaxation, breathing practice, mental imagery, and mindfulness, etc., which can help and accelerate practitioner access to meditative states”25. Members of the experimental group performed IBMT training for twenty minutes a day for five days, and, compared to

23 Ibid., p. 55.
the control group, significantly improved their scores on the Attention Network Test, a test that “measures skill in the resolution of mental conflict induced by competing stimuli”\textsuperscript{26} by, for instance, having participants respond to the direction of a random series of arrows by hitting the corresponding arrow key, but with the target arrow surrounded by others that randomly alternate between either pointing the same direction of the target arrow or facing a different direction, testing both the ability to positively attend to the target arrow and to keep one’s response from being affected by paying attention to the distracters. Participants in this study also showed “lower anxiety, depression, anger, and fatigue, and higher vigor on the Profile of Mood States scale, a significant decrease in stress-related cortisol, and an increase in immunoreactivity.”\textsuperscript{27} The further results for participants’ emotional states in these studies, of course, suggests a link between improved attentional capacities and more positive affective states. As I will argue in 5.2.4, this link gives us reason to think that training can help shape our emotions.

A further reason to think that rituals can serve as a form of attentional training that will yield results for our attentional tendencies comes from the work of Martin Seligman, an advocate for positive psychology. Seligman tested the effectiveness of an intervention that he has, at various times, called the “what went well”, “three blessings”, and “three good things”\textsuperscript{28} exercise. This exercise involves a daily ritual of coming up with three things that went well over the course of the day, and writing them down along with an explanation of why they went well. According to Seligman, this ritual helps us to “get better at thinking

\textsuperscript{26} Ibid., p. 17153.
\textsuperscript{27} Ibid., p. 17152.
\textsuperscript{28} Martin Seligman, \textit{Flourish: The New Science of Positive Psychology} (New York: Free Press, 2011) For each use of these names to refer to nearly identical exercises, see p. 33 and p. 84.
about and savoring what went well” as a corrective to our natural tendency to more vividly focus on the negative aspects of our lives, i.e. what went wrong, what problems are still to be resolved, etc. According to Seligman,

“Participants in the *three good things* exercise began to show beneficial effects at one month following the post-test. At the one month follow-up, participants in this exercise were happier and less depressed than they had been at baseline, and they stayed happier and less depressed at the three month and six month follow-ups.”

How did the exercise yield this result? It seems to me that it does so largely by shaping one’s attentional tendencies. By performing the exercise, one practices looking at the positives in one’s life. The daily nature of the exercise helps to make it the focus of one’s day, as everything that one does is leading toward the “what went well” exercise. This, over time, draws one’s attention throughout the day to what is happening that might end up on the list that night. This keeps one’s focus on the good in one’s life, not just at the point of performing the exercise, but as a matter of course. This helps to keep thought patterns from spiraling toward the negative, and creates a sort of pro-positive bias in one’s daily course of evaluative perception (seeing things as good, bad, etc.). On particularly bad days, one may have to work to find the good. Consistently attempting to do so also opens us up to noticing good things throughout one’s day that may have escaped notice otherwise, or to find the good in an upsetting situation more effectively in real time. In addition to being a

---

29 Ibid. p. 33.
30 Martin E. P. Seligman, Tracy A. Steen, Nansook Park and Christopher Peterson, "Positive Psychology Progress: Empirical Validation of Interventions," *The American Psychologist* 60, no. 5 (2005): 410-421. Regarding the effectiveness of the intervention in producing the result, Seligman also says, “There was a significant effect for adherence to the exercise on happiness scores at all time periods and a significant effect for adherence to the exercise on depression scores at the one-month follow-up point. The interaction between continuing the exercise and adherence to the exercise was significant for happiness scores, indicating that participants who continued the exercises were the happiest.” See p. 419.
plausible story, I believe that this explanation comports with the results I discussed above that show that with greater attentional control comes more positive affective states and a diminished stress response. This can perhaps lead to a compounding effect, as the diminished stress response may give people further access to greater attentional control, simply by removing distracting internal stimuli.31

A system of rituals, then, in which we are to be aware of our emotions, our comportment, and our norm-laden roles, or in which we otherwise reflect on our conduct throughout the day like in Zengzi’s daily examination of himself that I mentioned in Chapter 432, is likely to be effective through the same sorts of processes. The ritualized attending makes certain morally relevant aspects of our inner and outer experiences more salient to us in the same way that Seligman’s ritual makes the positive aspects of these experiences salient to us.

Is the increased salience of that to which Confucian rituals draw our attention capable of having a beneficial, pro-moral effect on our behavior, though? In other words, is there support for claim (2)? I think we have plenty of reasons to believe that there is. For starters, many of the results in the situationist literature that show differences in people’s behavior appear to turn on the degree to which their situations make morally relevant factors of their environments salient to them. For instance, one study manipulated different forms of self-directed attention to see the effects that this had on helping behavior, and found that

---

31 As I will discuss in 5.2.4, this feedback loop between attentional control and affect may be part of what is behind the correlation between positive affect and helping behavior, with positive affect giving us a cognitive space in which it is easier to attend to the parts of the world that cue us to help others.

32 Analects 1.4. Zengzi said: “Daily I examine my person on three counts. In my undertakings with people, have I failed to do my utmost? In my interactions with colleagues and friends, have I failed to make good on my word? In what has been passed on to me, have I failed to carry it into practice?”
certain conditions needed to obtain in order for self-directed attention to have a positive
effect on helping behavior rather than a negative one: “The situation must clearly set off an
orientation toward acting on a value of helping; that is, the cue to help must be legitimate as
well as salient [and the] person who is called upon to act pro-socially must not come into the
helping situation with a personal preoccupation that would be inimical to thinking about
helping.”\(^3\)\(^3\) The mindfulness skills that people can build through Confucian ritual training
help us to meet these conditions in a couple of ways. First, by strengthening people’s
attentional capacities, they allow people to better and more consistently attend to their
experiences, so as to allow both helping opportunities and their helping motives to be more
salient. Second, through the aforementioned reduction in stress and ruminative behaviors,
they free people of the sorts of personal preoccupations that are inimical to thinking about
helping.

In addition to building attentional skills and helping to change our thought patterns,
ritualized action also appears to help draw our attention to certain aspects of our situations,
and to keep us from being distracted by our personal preoccupations, in real time. A recent
study showed that when people performed ritualized actions before eating food, they found
the food more “flavorful, valuable, and worthy of consumption.”\(^3\)\(^4\) This effect was only
found when the action was ritualized rather than random, and when the person was
participating in the ritual rather than merely observing it, which recalls the discussion from
Chapter 4 that a key element of Confucian ritual training involves cultivating a sense of

---

\(^3\) Fredreck X. Gibbons and Robert A. Wicklund, "Self-focused Attention and Helping Behavior," *Journal of

\(^4\) Kathleen D. Vohs, Yajin Wang, Francesca Gino and Michael I. Norton, "Rituals Enhance Consumption,"
personal presence in the ritual so as to allow it to be more effective. The authors of the study also found that a delay between the ritual and the eating boosted the effect, which “attests to the idea that ritual behavior stimulates goal-directed action”\textsuperscript{35}, drawing one’s attention to goal-specific aspects of one’s experience in a way also allows for a “greater involvement in the experience.”\textsuperscript{36} If we need to attend to the world in a certain way in order to navigate it with optimal moral success, and rituals can have this effect of stimulating goal-directed action and creating a sense of presence that allows for full attention, then it appears that rituals, in addition to being used as a form of attentional training, can be effectively placed throughout our lives and social interactions so as to keep our eyes on the moral prize.

We can further see the effect of attention on behavior in Darley and Batson’s famous good Samaritan experiment\textsuperscript{37}, in which being hurried had a strong effect on whether seminary students stopped to help a person who was slumped over and apparently in need of assistance, can be plausibly interpreted as showing that different degrees of salience have an effect on whether or not people exhibit helping behavior. A slower-walking student was able to attend more clearly to the person in need than a hurried one, and this led slower students to help more often. The hurried students were focusing more clearly on getting to their destination and exerting the intentional effort that it took to hurry, and so they exhibited a typical case of inattentional blindness\textsuperscript{38} that led them to not really notice the person in need, or to attend to the person enough to make the connection that the person

\textsuperscript{35} Ibid.

\textsuperscript{36} Ibid.


was in need. Given the real-time effects of rituals on attention, perhaps if these students had developed even so much as a personal, yet conscious, ritual way of greeting strangers whenever they were on campus, the ritual may have drawn their attention to the person’s need by activating a goal of, say, sincerely inquiring as to how the stranger is doing, and perhaps this would have made more salient the fact that the person was in need and induced more helping behavior even in the hurried condition.

So, attention is something that we appear to be able to train through mindfulness techniques that are an important part of Confucian ritual training, and proper, goal-directed attending is something that rituals can help to activate within us in real time. Where and how we direct our attention also seems to be behaviorally relevant. Because of the connection between attentional training, executive functions, and moods, a good deal of the evidence that I cite in subsequent sections will further support the link between attentional training and behavior, but this should add to what, for now, already appears to be solid support for claims (1) and (2).

5.2.2 Executive Control – Claims (3) and (4)

The first thing to note as I move forward to defend claims (3) and (4), i.e. that our executive functions can be strengthened through training by ritual activities and that strengthening these functions can improve our ability to comply with our moral norms, is that the ability to direct and sustain one’s attention is itself one application of our executive functions, which includes other skills like emotion regulation and behavioral control. A
recent neurological study on adults with Attention Deficit-Hyperactivity Disorder showed a biological link between attentional capacity and executive function by finding

“significant cortical thinning in ADHD in a distinct cortical network supporting attention especially in the right hemisphere involving the inferior parietal lobule, the dorsolateral prefrontal, and the anterior cingulate cortices. This is the first documentation that ADHD in adults is associated with thinner cortex in the cortical networks that modulate attention and EF [(executive functions)].”

Given this link between attentional capacity and executive function, it is likely that the arguments in section 5.2.1 already serve as some evidence for the claims that I am defending in this section, since greater or lesser attentional capacity appears to be correlated with greater or lesser executive function.

This connection between executive cognition and attention can also often be seen in the way in which psychologist operationalize and test executive functions. A recent study hypothesized “that mindfulness meditation practice aids the development of two facets of executive cognition,” those facets being our ability to sustain attention on a particular stimulus and our ability to volitionally switch attention from one stimulus to another. These abilities to sustain and switch attention are, essentially, abilities of executive self-regulation, which explains why recent studies on mindfulness suggest that “attentional control may lead to increased self-regulation.” To exercise our ability to keep our attention in one place is to exercise our general capacity for volitional control, i.e. a capacity to determine what we do and do not do. Depending on how strong of neural overlap there actually is between our

---

39 Nikos Makris, et al., "Cortical Thinning of the Attention and Executive Function Networks in Adults with Attention-deficit/Hyperactivity Disorder," *Cerebral Cortex* 17, no. 6 (2007), p. 1364.
41 Ibid., p. 306.
capacity to control our attention and what we might think of as a conceptually more-general capacity for volitional control, we may want to say either that our attentional control capacities underlie our executive control capacities, that our attentional capacities simply are our executive control capacities, or that strengthening a general capacity for volitional control in turn improves our capacity to determine the sorts of behaviorally-relevant stimuli, both internal and external, to which we subject ourselves; either way, their link is undeniable.

For instance, somebody who ruminates about being slighted rather than noticing the rumination and then directing one’s attention elsewhere spends more time attending to and feeling negative emotions like anger and resentment, and greater attention to these emotions may lead to a greater likelihood that one engages in a retributive behavior. Our executive functions not only allow us to withstand certain behavioral impulses; they also allow us to choose what is present to our minds, which helps to determine the nature of the impulses that are there to be managed or overcome in the first place. This is perhaps why, in the study I cited in the previous paragraph, it was found that “those completing the mindfulness training demonstrated significant improvements in self-reported mindfulness, depressive symptoms, rumination, and performance measures of working memory and sustained attention”; this complex of improvements is most likely interrelated, because improving attention and working memory enables one to better hold certain thoughts in one’s mind to the exclusion of other thoughts, essentially crowding out ruminative cognitive behaviors and negatively-valenced thoughts.

While, in general, as I have discussed, training that improves the function of the circuitry that supports executive cognition and working memory has been shown to improve
emotion regulation, it has recently been shown that working memory tasks that deal with specifically emotional content have an especially significant effect on both brain function and emotion regulation capacity. One standard working memory task presents participants with a series of images, symbols, words, etc. and has them respond differentially if the current image/symbol/word corresponds to the image/symbol/word either one (easy), two (medium), or three (hard) places back. For instance, if the current image is a triangle, they would hit the agreement button if they were presented with a triangle three images ago, and the disagreement button if it was something else. In the emotional working memory task, the images presented were either faces of characteristic emotions or emotionally-laden words. Training on this kind of emotionally-laden working memory task “enhanced the efficiency of this frontoparietal demand network. Critically, compared with placebo training, emotional WM training also accrued transfer benefits to a ‘gold standard’ measure of affective cognitive control – emotion regulation.”42 The increased emotion regulation capacity was “associated with greater activity in the targeted frontoparietal demand network along with other brain regions implicated in affective control, notably the subgenual anterior cingulate cortex.”43 This helps to show that training that is emotion-specific efficiently trains the capacities that we need to train in order to have better top-down emotional control. Since, as I have discussed, so much of Confucian ritual training involves focus on aspects of one’s inner and outer experience that is emotionally relevant, like what the emotional content of one’s face or voice should be on a given occasion or what emotions are being

---

43 Ibid.
evidenced by other people’s mannerisms, this result for emotion-specific training suggests that Confucian rituals are likely effective at inculcating affective control.

Further evidence for this view, that greater executive functions can work in concert with a greater capacity to direct one’s attention to make choosing the right behaviors less difficult, comes from Walter Mischel’s tests on delayed gratification. In this study, children were presented with a treat, such as a marshmallow, and told that they could have a larger treat if they could wait 15 minutes without eating the initial treat. Executive control of attention was a significant determinant of success on the task, as “those children who are able to direct their attention away from the reward-related stimuli in the task (low temptation focus) are able to wait longer than those children who direct their attention toward the reward-related stimuli (high temptation focus).”\(^44\) That this is either the result of, or comes part and parcel with, an underlying executive capacity can be seen by a demonstrated correlation between success of the delayed gratification task and go/no-go task\(^45\), which tests response inhibition by presenting subjects with a continuous stream of stimuli and instructs them to respond to one kind of stimulus by hitting a button (go), but to avoid hitting the button when presented with a different kind of stimulus (no-go). The “go” stimulus is typically more frequent, leading success in the no-go event to require inhibition of a prepotent response to hit the button.

Support for the link between directed attention and a more general executive capacity like response inhibition is bolstered by the fact that “circumstantial evidence

\(^{44}\) Inge-Marie Eigsti, et al., "Predicting Cognitive Control From Preschool to Late Adolescence and Young Adulthood," *Psychological Science* 17, no. 6 (2006), p. 479. This quote is discussing Walter Mischel, Yuichi Shoda and Monica L. Rodriguez, "Delay of Gratification in Children,," *Science* 244 (1989): 933-938.

\(^{45}\) Ibid., p. 478.
indicates that performance in the delay-of-gratification and go/no-go tasks [(a response inhibition task)] reflects similar biological and neural systems. The link to behavior is also rather robust, as longitudinal studies have shown that those capable of delaying gratification as children “have been found to be more attentive, to be better able to concentrate, and to exhibit greater self-control and frustration tolerance than their peers when they are adolescents…score higher on the SATs and are perceived as more interpersonally competent by parents and peers [and as] adults, they are less likely to use drugs.” So, executive functions and the ability to attend to the right sorts of stimuli are both connected to each other and behaviorally relevant.

It has also been shown that mindfulness meditation can serve as a real-time strategy for helping people to overcome the effects of self-control depletion. Participants in one study had their conscious control depleted through an exercise in which they were told to suppress their emotional reactions while being shown videos intended to elicit disgust, including videos of medical procedures and “a close-up of someone squeezing a massive pimple on another person’s back.” They were then presented with a subsequent task in which they had to scan sheets of paper filled with “d”s and “p”s that were each surrounded by one to four apostrophes, and were told to quickly cross out all and only the “d”s with 2

46 Ibid., p.479.
apostrophes, a task with has been “successfully used previously as a measure of self-control strength.” Some participants were instructed to engage in mindfulness meditation between these tasks, and the study found that “participants who had meditated after emotion suppression performed equally well on the subsequent self-control task as participants who had not exerted self-control previously.” This result suggests that people whose daily lives are permeated with ritual moments that call for and allow for the kind of mindfulness I have discussed will have a greater degree of conscious control over their behavior throughout the day, not only through strengthening those conscious control resources, but by having moments each day to replenish those resources. Recall, as I discussed in Chapter 4, that this sort of permeation of ritualized activity throughout one’s life, from prescribed ways to perform daily activities like eating to prescribed social rituals that one can use throughout one’s social interactions, is a hallmark of the Confucian strategy. So, it seems as though we have good reason to accept claims (3) and (4).

5.2.3 Emotional Insight – Claims (5) and (6)

The capacity for emotional insight, both intrapersonally and interpersonally, helps one to more consistently perform morally correct behaviors. The ability to recognize and understand one’s emotions and moods in real time can help keep people from being unduly influenced by emotions or mood effects that would lead them astray, by affording them the opportunity to detach from emotions like anger and anxiety. This intrapersonal emotional

\[\text{Ibid. p. 1019.}\]
\[\text{Ibid., p. 1016.}\]
insight also affords one the opportunity to choose to identify with positive, pro-social emotions when they occur, helping to make them more salient and thus more likely to induce helping behavior, generous behavior, and so on. Interpersonal emotional insight also aids our capacity for moral behavior by allowing us to understand other people’s emotional states, which helps tell us how we ought to treat them and provides opportunities for drawing motivation from sympathy. By making us pay attention to our bodily states, facial expressions, comportment, etc., and by prescribing particular expressions of emotion in particular circumstances, Confucian rituals can help to inculcate that capacity for emotional insight.

Much of the evidence that I have already discussed in section 5.2.1 supports the idea that Confucian ritual training can improve people's abilities at recognizing their own emotions. The results that mindfulness programs have gotten in reducing negative affect and improving certain behavioral problems come in large part from the improving people’s ability to recognize and detach from one’s emotions. This is part of the reason why mindfulness-based interventions are effective in the treatment of borderline personality disorder, which has been associated with a lowered ability to represent one’s own emotional experiences51, a lowered ability to recognize the emotional content of other’s faces, as well as a seemingly related intensity of negative affect.52 Education in skills in recognizing one’s own emotions are also a part of social and emotional learning programs that have been used

---


in schools to significant effect; a recent meta-analysis of studies on social and emotional learning programs found that they “yielded significant positive effects on targeted social-emotional competencies and attitudes about self, others, and school. They also enhanced students’ behavioral adjustment in the form of increased prosocial behaviors and reduced conduct and internalizing problems.”53 A variety of programs for teaching people how to notice, understand, categorize, and deal with their own emotions have been successful, and the Confucian ritual-based approach appears to share many of the same proximal goals as these interventions.

One further piece of evidence that an emotional training program can improve our emotional insight comes from a study in which participants engaged in an intensive 8-week course in mindfulness/emotion regulation that included training in how to recognize microexpressions. It was found that “Training participants increased their ability to recognize these subtle facial expressions of emotion, while the control group did not.”54 This increased ability in assessing the microexpressions of others was correlated, in this study, with a decrease in rumination, suggesting that the ability to better distinguish emotional expressions in the faces of others can improve one’s ability to notice and understand one’s own emotional experiences so as to keep oneself from engaging in cycles of thought that are biased by negative emotions.

One of the ways in which this training to recognize emotions in the faces of others works is via a type of attentional training. Learning to better discriminate between different

---

expressions of emotion means, in part, developing a habit to attend to the correct facial features during an expression of emotion. In one study on the effects of facial expression training, different gaze-tendencies were developed through training that improved the ability to detect emotions, and the tendencies that were strengthened appeared to be the tendencies that we build through typical emotional learning:

“Adults focused most on the eyes in all sessions and increased expression categorization accuracy after training coincided with a strengthening of this eye-bias in gaze allocation. In children, training-related behavioural improvements coincided with an overall shift in gaze-focus towards the eyes (resulting in more adult-like gaze-distributions) and towards the mouth for happy faces in the second fixation.”

This type of attention is an external counterpart of the kind of internal awareness that I have already discussed in section 5.2.1.

Work in the emotional training of children with autism has further shown the ways in which emotional training can improve interpersonal emotional insight. One training program used facial recognition software that was integrated into a game called “FaceMaze”, in which participants needed to make either “happy” or “angry” faces in order to proceed. The training was effective in helping the participants learn to express more characteristic and recognizable expressions of emotion: “Naïve raters judged the post-FaceMaze ‘happy’ and ‘angry’ expressions of the [Autism Spectrum Disorder] group as higher in quality than their pre-FaceMaze productions. Moreover, the post-game expressions of the ASD group were rated as equal in quality as the expressions of the [typically developing] group.” This means

---

that training one’s facial expressions can work as a form of emotional training. Training in properly performing certain emotional expressions should also improve one’s ability to recognize both emotions in oneself and emotions in others, given that facial mimicry is one of the avenues through which people understand and recognize the emotions of others. Collectively, this gives us reason to accept claims (5) and (6).

5.2.4 Shaping the Emotions – Claims (7) and (8)

I have already introduced some evidence, in section 5.2.1, that supports claims (7) and (8), i.e. that we can alter the content and/or strength of our emotional responses and moods through ritual activities, and that doing so can improve our ability to comply with our moral norms. To summarize, recall that attentional training has the effect of reducing anxiety, depression, anger, and fatigue and that positive psychology has yielded similar results through a daily ritual that draws our attention to the positive aspects of our life. Recall also that this sort of training has been shown to strengthen our executive functions, which can be key in making decisions to move our focus away from various negative emotional stimuli or thoughts, and toward more productive activities or cognitions, and that this has been shown to effect both moods and behaviors. This all serves to support the claim that rituals can help to shape and maintain our moods and occurrent emotions in ways that should serve to support pro-social, morally correct behavior. What I will present in this section is aimed at supplementing the relevant evidence that I have already discussed.

The study I cited in section 5.2.3, in which participants went through an intensive 8 week course in mindfulness meditation, also showed a link between the decreased negative affect that emotion-monitoring skills can engender and a greater likelihood of engaging in pro-social behavior. Not only did the group in the training condition report “reduced trait negative affect, rumination, depression, and anxiety, and increased trait positive affect and mindfulness compared to the control group,” as is consistent with other findings about mindfulness practices, but they also showed that the training “protected trainees from some of the psychophysiological effects” of a stress-inducing activity which involved having to develop a present a speech and do mental arithmetic in front of evaluators they had not met, “appeared to activate cognitive networks associated with compassion,” and reduced hostile behavior compared to controls in a task in which participants and their significant others spent fifteen minutes discussing a problem in their relationship. The study also showed that 5 months after training, many of the effects, such as decreased negative affect and decreased anxiety, were maintained. These results, especially the result that training increases the cognitive, motivational, and behavioral aspects of compassion, helps to show that the sort of training that we can get through Confucian rituals can significantly alter our make-up in ways that change our emotional responses, and that changing our emotional responses can make it easier to steer ourselves toward moral behavior. That we can improve behavior by altering people’s emotional states has also been shown outside of the context of behavioral therapies that rely on a mindfulness-based approach to one’s emotional life. Of course, it is

---

59 Ibid.
60 Ibid.
commonplace that negative emotions like anger and fear can lead to behavior that harms others or otherwise goes morally astray. It is also a robust result in the psychological literature, though, that positive affect, for instance, correlates with helping behavior and generosity.

As I discussed in section 5.2.3, the focus in Confucian rituals on exhibiting the right kind of facial expressions means that they can help one, though practice, to develop emotional insight; this facet of Confucian ritual practice can also, however, serve as an emotional regulatory tool. It has been shown that by consciously manipulating our facial expressions, we can alter our emotional experiences in two ways. First, manipulation of our faces into expressions that are characteristic of certain emotions has been shown to induce experience of those emotions. For instance, when study participants perform the Directed Facial Action task, in which they are given a set of individual instructions to move parts of their face until their faces are arranged such that they exhibit a characteristic expression of emotion, they tend to show both physiological changes characteristic of experiencing those emotions and also self-report feeling those emotions in ways that “are generalizable in pattern across age, profession, gender, and culture…[and] are similar to those produced using more ‘conventional’ emotion-eliciting tasks such as imagery, films, and situational manipulations.”

Other research has tended to support the Facial Feedback Hypothesis, the

---

claim that there is a feedback loop between our facial expressions and our experiences of emotion that allows manipulations of people’s facial expressions to affect the content of their emotional experiences. In one study, people were told to hold pencils in their mouths in certain ways, either inhibiting or eliciting various types of smiles. Compared to others, people who were made to display a genuine, “Duchenne” smile in this condition “reported more positive experience when pleasant scenes and humorous cartoons were [presented and] tended to exhibit different patterns of autonomic arousal when viewing positive scenes.”

What this all means is that we can actually induce certain emotional experiences by way of attending to our facial expressions; this should allow us to practice and perfect them through the performance of rituals by paying attention both to what our faces are doing and to the content of our emotional experiences as a result. We can learn to be more present in an emotionally-laden ritual by building a performance habit that better induces the appropriate emotion and the correlative physiological response. This also should work to instill various emotional habits, as certain facial expressions, and their attendant emotional state, become second nature.

The second way that ritualized episodes of attending to our facial expressions can help shape our emotions is by giving us an easy way of suppressing morally problematic emotions. It has been shown that focusing on keeping our faces still is an effective method of keeping ourselves from feeling certain emotions as strongly. One recent study found that participants who were told to keep their faces still, not as an emotional suppression task but

---

because they were hooked to electrodes that were purported to record information about attention and memory, reported significantly weaker emotions as the result of watching emotionally-laden video clips than those given no instructions, those wearing no electrodes, or those told to distract themselves through the clip by counting backward in threes from one thousand. This further suggests that the focus on one’s countenance that Confucian rituals can teach, rehearse, and instill can be used as an emotion regulation strategy; if a situation elicits a negative emotion like anger or resentment, a focus on one’s face can weaken the effect of that emotion and allow one to overcome it and potentially choose a morally correct behavior over an incorrect one. It also then appears likely that one can build a habit of responding to certain negative-emotion eliciting situations by being mindful of one’s face, so as to keep negative emotions at bay.

Given the potential for these methods for determining the content of our emotions, as well as the general emotional benefits of mindfulness, it appears that we have good reason to accept claim (7). That we also have reason to accept claim (8) can be seen, first, from the various connections we saw in section 5.2.1 between the emotional and behavioral effects of mindfulness practices; the behavioral effects of mindfulness appear to come part and parcel with the emotional effects. These results also, of course, accord with our commonsense understanding of the way in which emotions drive behavior; when we reduce the degree to which people feel emotions such as anger and anxiety, we reduce the opportunity for these emotions to have their often negative behavioral effects. The ability to either induce or suppress certain emotions via rituals or the physical-emotional biofeedback skills we develop

---

from rituals can also serve an intervening role similar to executive control, interrupting lines of emotionally-laden thought that can lead to unacceptable behavior. Furthermore, studies have shown that inducing positive affect can lead to pro-social behavior, even if the induction of positive affect is the result of something as minor as finding a dime or being given cookies, or because the room that people were in was filled with a pleasant fragrance. This is in part because, as I discussed in section 5.2.1, positive affect and the resultant reduction in negative thought spirals that comes along with more consistent positive affect makes it easier to attend to the world, including to the morally relevant aspects of the world that we need to notice in order to be morally successful. So, claims (7) and (8) appear to be well justified.

5.2.5 Roles and Role Attributions – Claims (9) and (10)

That social roles have an effect on our behavior is a well-worn notion in both sociology and social psychology, and has been studied under the label of “role theory”. Role theory “may be said to concern itself with a triad of concepts: patterned and characteristic social behaviors, parts or identities that are assumed by social participants, and scripts or expectations that are understood by all and adhered to by performers.” The basic idea is that we use role terms, like “mother”, “teacher”, “friend”, “sick person” and so on, as a way

---

of situating ourselves and each other in a nexus of behavioral expectations that influence our self-concepts and motivations in ways that ultimately affect our behavior.

Social role theory has been used to explain particular behavior patterns, and is connected with labeling theory, the idea that the particular labels that we bestow upon people affect their behavior. This can happen in both negative and positive ways. A group of studies on the negative impact of labels have focused on the effects of labeling juveniles as delinquents, and the effect that this has on their behavior. One study found that “parents’ labeling and youths’ perceptions of teachers’ labeling significantly and positively related to subsequent delinquency.”69 In a study on the effects of being officially labeled a “felon” on criminal behavior in adults, “data for 95,919 men and women who were either adjudicated or had adjudication withheld show that those formally labeled are significantly more likely to recidivate in 2 years than those who are not.”70 How much this finding has to do with the direct effect of the label on self-concept is unclear, because the “felon” label carries with it its own set of burdens and social stigma, but even this helps to highlight the need for us to have the right kinds of role terms. The results of another study on incarcerated youths “provided partial support for the hypothesis that juveniles who choose a greater number of negative labels for their self-concepts will report more frequent delinquent involvement. Labeling variables were significant predictors of general and serious delinquency.”71 That latter finding more clearly illuminates the relationship between labeling, self-concept, and

---

behavior that gives labeling the power that it has. Labels help to tell us who we are, and we act, at least in part, in accordance with our conception of who we are.\textsuperscript{72}

Labeling like this has also been shown to influence behavior in positive, pro-moral directions, and to be more effective than exhortations to better behavior. In one famous study, students who were given an award for tidiness, and were reminded that they were a tidy class, ended up displaying tidy behavior more reliably and for longer after the intervention than students who were given a lesson on the need to be tidy.\textsuperscript{73} Another study found a similar result, namely that arguments in favor of environmentally-friendly behaviors were less effective at inducing environmentally friendly choices than labeling people as environmentally friendly. While it has not been studied how effective role labeling is in comparison with trait labeling in shaping our self-concepts in behaviorally-relevant ways, I believe these studies on trait labeling at least serve to indirectly support the notion that role labeling has effects on people’s behavior, because role terms are essentially terms that stand in for a bundle of traits, expectations, etc. While there is some difference between calling somebody a “caretaker” and calling them “caring” or between calling somebody a mother and calling them “motherly”, in that the former terms more directly convey the normative expectation than the judgment of normative success and the latter terms do the opposite, it

\textsuperscript{72} Cf. Jesse Prinz, "The Normativity Challenge: Why the Empirical Reality of Traits Will Not Save Virtue Ethics," \textit{The Journal of Ethics} 12 (2009): 117-144. wherein he says “If you see yourself as a hippy, you might wear certain clothing and advocate particular causes. Your preferences would be different if you defined yourself as a redneck. If you see yourself as gay, or as a Jew, or as preppy, you might adopt patterns of behavior that associated with each of those categories. We often live in accordance with prescribed social roles, and, when we do, we often see these roles as expressions of our identity—they are part of our character, as that term is ordinarily used.”

strikes me that a similar enough attribution is being made for us to think that a similar 
labeling effect will occur.\(^{74}\)

One powerful piece of evidence for the claim that being attributed a certain role can 
have an effect on people’s behavior that we would do well to harness and point in the right 
direction comes from the Stanford Prison Experiment that I discussed in Chapter 2.\(^{75}\) In 
this experiment, students were randomly assigned to the roles of either a prisoner or a prison 
guard, and their behavior on each side reflected these roles in ways that led to a number of 
ethically problematic behaviors. Recall that criticism of the experiment came in the form of 
denying that it was an ecologically valid look at the power dynamics of actual prisons and the 
behavior that they might produce. Psychologist Peter Gray offers this alternative 

explanation of the participants’ unusually immoral behavior:

> “Twenty-one boys (OK, young men) are asked to play a game of prisoners and 
guards. It’s 1971. There have recently been many news reports about prison riots 
and the brutality of guards. So, in this game, what are these young men supposed to 
do? Are they supposed to sit around talking pleasantly with one another about 

sports, girlfriends, movies, and such? No, of course not. This is a study of prisoners 
and guards, so their job clearly is to act like prisoners and guards—or, more 
accurately, to act out their stereotyped views of what prisoners and guards do.”\(^{76}\)

So, the experiment, to Gray, shows not that it was being placed in a certain situation that had 
the effect on the participants’ behavior, but rather that it was taking on a social role had that 
effect. He elaborates that the expectations that they perform their roles correctly for the 

experimenters influenced them to comply with the roles even when they could have walked

\(^ {74}\) I thank David Wong for bringing to my attention this way of explaining the difference between what is conveyed by role terms and what is conveyed by trait terms.


\(^ {76}\) Peter Gray, "'Why Zimbardo's Prison Experiment Isn't in My Textbook: The Results of the Famous Stanford Prison Experiment Have a Trivial Explanation," *Psychology Today*, October 9, 2013.
away, which, the way I interpret it, meant that there was an expectation that mimicked what happens in the real world when people are expressly situated in particular roles. To whatever degree trait labeling and role labeling differ, then, it does not appear that the greater focus on normative expectations that we see in role terms make them any less effective at inducing label-consonant behaviors.

In addition to this explicit assigning of roles, simply being reminded of particular roles through the use of framing effects and priming effects has been shown to have an influence on our behavior. In studies in behavioral economics, for instance, it has been shown that giving different labels to a particular game, labels that influence how participants see their roles as players of the game, influences people’s willingness to act in pro-social ways. For instance, one study that looked at the effectiveness of labeling versus reputation on cooperative behavior had different groups of participants play a prisoner’s dilemma game. For one group, the game was called the “Wall Street Game”, and for the other it was called the “Community Game”. As it turned out, “[when] playing the Community Game, 67% of the most likely to cooperate nominees and 75% of the most likely to defect nominees cooperated on the first round. When playing the Wall Street Game, 33% of participants with each nomination status cooperated”.\(^7\) I think a plausible explanation of this result is that players of the Wall Street Game are primed to conceive of their roles in a way that includes maximizing personal profits, outmaneuvering others, and so on, whereas

players of the Community Game see their role as that of a “community member”, which means being cooperative, pro-social, and the like.

Further evidence for the behavioral efficacy of roles and role terms can be found in the work on the effects of stereotypes. With broad brush strokes, we associate certain qualities with certain groups of people, whether they are demographic groups like “eldery” or “African American” or social roles like “professor” or “politician”. There are various ways in which we can activate these stereotypes out in the world such that they have a behavioral effect, such as conscious or unconscious priming. Sometimes, the activation of a stereotype is tied to one’s self-concept. For instance, the literature shows a robust phenomenon known as stereotype threat, in which “When aspects of one's identity related to task performance are made salient, performance is affected in the direction of that aspect of identity.”

For instance, women, stereotyped as being less good at math than men, did worse on math tests when their gender was made salient, the elderly did worse on memory tests when the negative aspects of being elderly was made salient and did better when the positive aspects of being elderly were made salient, and Asian Americans, stereotyped as being adept at math, did better on math tests when their race was made more salient.

In other cases, the stereotype that is activated need not be related to one’s self...

---

concept. In another study, for instance, after showing that politicians are pretty reliably stereotyped as being long winded, experimenters primed people to think of politicians by way of a word-unscramble task. When they were then asked to write essays on a foreign policy matter, the people who were primed wrote much longer essays, showing that the stereotype had been activated when people did not identify with them. This result has been borne out through other studies showing that people move more slowly or more quickly depending on whether and in what way they were primed with either stereotypically fast or slow animals, and that even young people move more slowly when they have been primed with the idea of elderliness. Interestingly, though, it has also been shown that the effect of stereotype activation is stronger in cases in which people identify with the stereotype, even if it does not actually apply to them. For instance, in one study, the negative stereotypes of African-Americans were activated by way of an instruction to write about a day in the life of an African American, and people who chose to write in the first person showed a greater stereotype effect than those who chose to write in the third person, indicating that stereotype activation and stereotype identification are probably distinct effects.

The fact that we find effects both in activating stereotypes that are a part of somebody’s self-concept, like in stereotype threat, and in priming people by way of existing

---

stereotypes that they have but with which they might not identify, like in the animal, professor, and elderly cases, helps to support the idea that the Confucian role attribution strategies that I laid out in Chapter 4 should be effective. Recall, those strategies included both situating ourselves in well-defined roles with which we identify, and then using “off-label” role ascriptions in contexts in which the social roles do not actually obtain, like activating the stereotype for “son” in one’s performance of duties to one’s superiors.

In addition to role attributions being effective in both the self-concept and non-self-concept conditions, we have reason to believe that a mechanism like rituals, by giving us consistent experiences that we associate with role terms, can help to make them effective in the ways we want them to be; names, in other words, can be rectified. It has recently been shown that our stereotypes are malleable via different informational inputs. One study on the stereotyping of groups “collected participants' beliefs about the occupational roles (e.g., lawyer, teacher, fast food worker, chief executive officer…) in which members of social groups (e.g., Black women, Hispanics, White men, the rich, senior citizens, high school dropouts) are overrepresented relative to their numbers in the general population” and found the sorts of stereotype effects that have been typical in the literature. One experimental group was then presented with new, projective descriptions of the social groups that included changes in their future occupational roles, after which it was found that the subsequent stereotype effects “were more influenced by these future roles than by their current group stereotypes, thus supporting social role theory’s predictions about stereotype

---

change. 

This indicates not only that projective social role attributions or new social role enactments can alter our stereotyped views of people, which is an effect that can be put to good moral use, but it also suggests that, insofar as role terms are themselves vehicles for behaviorally-effective stereotypes, that it is possible to alter how those stereotypes work by updating the content of the stereotypes. This supports the idea that, because they situate us in our roles and give us experiences that help to shape our background understanding of role terms, Confucian rituals can be put to use in shaping the content of our role stereotypes and should thus, given the labeling and stereotyping effects that I have discussed in this section, should have an effect on our behavior that we can shape and maintain through the performance of role-laden rituals. This behavioral effect, of course, can also go beyond the automatic effects, like stereotype threat, by way of the mindfulness elements of the Confucian method, as ritual engagement with others out in the world can serve to draw our attention to our roles, and make the associated duties conscious and salient as we go throughout our days, giving us an opportunity to identify with and choose role-consonant actions.

Given the degree to which roles can effect our behavior via our self-concepts, stereotype activation, priming, and making certain aspects of our self-concept more salient to us, we would do well to follow the Confucian approach and proactively tend to these roles and role terms and the conditions in which we reinforce and activate them. Rituals can give us a way of both shaping the content of these roles and building their activation into our daily lives, giving us a way of putting the effects of roles to work for us in ways that can

---

87 Ibid.
positively effect the moral compliance of our behavior. We are thus, I believe, justified in accepting claims (9) and (10).

5.2.5.1 Rituals, Roles, and Social Contexts: Some Things to Note

Having shown that role attributions are an effective way of encouraging morally correct behavior, and that rituals can be used to activate and maintain these effects, I still need to contend with various complications that result from the social and cultural nature of both social rituals and the roles that they allow us to exercise and maintain. If the cultivation strategy involves rituals and roles that are so heavily social in nature, then proper implementation of the strategy is something that appears to require a proper social context, in which others buy into the strategy, agree both with the content of the rituals and in the commitment to perform them, and agree with the content of the social roles that are enacted and reinforced through these rituals. This may present a strong impediment to actually implementing the ritual strategy, especially in our increasingly cosmopolitan societies, or in societies that do not have any strong connections to traditional rituals or what we might call a “ritual infrastructure”. As such, I shall discuss, in this section, how we should understand how social context effects how we should think about implementing the Confucian strategy.

As a starting point for this discussion, we can note that much of what I attempted to show in sections 5.2.1-5.2.4 involved the effects of ritualized activity as mindfulness training tools, and in this context and for this purpose, the social aspect is relatively minimal. I can have a particular ritualized way of eating, a ritual for reflecting on my day, a ritual that I perform upon waking up, and so on, and I can garner all of the attentional, executive, and
emotional benefits of these activities. So, rituals can be, in at least some cases, individual in nature, and no less effective for it. Some rituals are perfectly effective and meaningful between two people, like best friends or spouses. Some rituals can be family rituals, or the rituals of a group like the girl scouts, the members of a church, the members of a drum and bugle corps, or the members of a philosophy department. Other rituals can be society-wide. To generalize this point, we can say that while rituals do need to be regularized to be effective, and that as rituals become more social in nature, this regularization must occur within an appropriately-sized social group, this can occur in groups of various sizes, from the individual potentially all the way up to an entire society.88

An initial practical upshot of noting this is that people who wish to implement the Confucian strategy have a number of available opportunities to try to do so in their own lives, in their friendships, in their families, in their workplaces, and in whatever social circles that they can convince of the benefits of imbuing their lives with rituals. This means that, for most intents and purposes, a person need not be in a society with a strong ritual infrastructure in order to put the strategy into effect. This is even true when the goal of rituals is the maintenance of certain role terms. If I am daily confronted with my role as a husband in the context of a family ritual, this meaning of this term and the effect of its attribution will be largely the result of the effects of those rituals, and so being reminded of my role, even among people who largely do not share my rituals, can still activate role-consonant conduct. Some role terms, like “boy scout” or “Christian”, are also heavily tied to a circumscribed social group, and yet can have beneficial effects on one’s behavior if the

88 Of course, as I discuss a little further on, the broader the social system we want the rituals to be regularized, the more difficult it is.
right kinds of rituals are used within that circumscribed group to imbued those terms with the right normative meanings.

This should give us hope that even in a pluralistic society where people come from various traditions or even in a society or social circles with no traditions at all, the Confucian strategy can be still implemented widely and to good effect through being implemented locally, as the vast majority of social interactions and opportunities for ritual conduct that we have are heavily local in scope anyway. In social circles where there are already rituals in place, people can reap moral benefits by paying attention to the effects of these rituals and making the most of their cultivatory potential. In social circles that are lacking rituals, people have the opportunity to design some, practice them, and maintain them, so that they too can reap their benefits. We should note, also, that because of the promise of setting up local rituals, pluralism does not present a problem in the way that it might if we needed to get some sort of massive social consensus about the right rituals as a prerequisite to getting the project off the ground. The more of that consensus that we can get, of course, the more potentially effective the Confucian strategy will be, because having broader consensus means having more and more social contexts in which we can implement consistent rituals.

Even if we accept what I have said thus far, are there not still ways in which cultural pluralism and pluralism about values can create conflicts that will upset the effectiveness of the Confucian strategy? For instance, are there not legitimate conflicts over what content certain social roles terms like “husband”, “mother”, “wife”, “citizen”, and so forth should have, and will this not lead to further conflict over the rituals that should be used to help maintain them? Here, I must admit that of course such conflicts exist; fortunately, I do not
think that it means much to the potential effectiveness of the Confucian strategy. First, with respect to terms like “husband” and “wife”, if different social groups have different conceptions of what exactly these terms should mean, or even if certain groups would prefer less gender-laden terms, then we do have disagreement, but so long as the ritualized attributions of them are, as I have discussed, mostly local in nature, then there is disagreement, but that disagreement should not very often lead to the kind of practical conflict that will upset the behaviorally-relevant attributions of those roles, or the system of rituals that keeps them meaningful and effective. It may mean that we cannot, as a larger society, develop rituals that uniformly maintain consensus about those terms, but local consensus is all we really need in these cases, especially if the goal is for people to consistently accord with the moral norms that they accept. My spouse rituals may not be your spouse rituals, but that does not keep each of us from similarly using rituals to be the best spouses we can by our own lights.

There are also different sorts of conflict, like when people simply will not play along. Perhaps somebody does not care to learn or properly perform a ritual, and when I greet this person on the street in the ritualized manner my community has developed, he simply shrugs it off. This does not seem to me to be a problem unless there are enough people like him in my community that the ritual simply breaks down, or does not gain a foothold. This would, of course, pose a problem for the Confucian method, since it would rob people of some of the opportunities to perform rituals and accrue their benefits. This, again, seems to me to be another case where the solution is to embrace the localization of rituals. While implementing the Confucian method requires people to situate themselves with cooperators
in the ritual project, it does not require them to be wholly surrounded by such cooperators. People simply need enough opportunities throughout their day to engage in rituals and experience their effects.

In cases like the term “citizen”, however, the role is one that every member of our society shares, and if we are to make the most of its moral effect, we should be able to design rituals that give this a particular meaning that we can broadly maintain as a society. Local rituals will seemingly not cut it in this case, because the role is one that must find its meaning in the much broader social system. Cases like this, in which the nature of the role itself requires much broader cooperation, may be one of the places were the disagreement about values can impede the implementation of the Confucian strategy. In these cases, however, there is at least the opportunity of negotiating some sort of consensus, and even if there is not, we may find that we do not need to make use of such broadly social role terms in order to have what we need to steer us in a consistently moral manner. Whether this is the case or whether we instead may need to find consensus on terms like these, remains to be seen.

With the perils of pluralism, however, also comes opportunity. Our pluralistic society, our globalized world, and the project of comparative ethics that both make possible, present us with an opportunity to draw from various traditions and have a conversation about which values we want, which practices make sense, and so on. Given the argument that I am presenting here in support of seeing the Confucian strategy as potentially effective, I believe that we ought to pursue that conversation about what values we want in ways that reflect the value and importance of discussing which conceptions of social roles we want,
what sorts of practices we ought to ritualize, how broadly we need to gain consensus on these matters, and where it makes more sense to allow for pluralism. There are many questions yet to answer in this conversation about the best way to implement the Confucian strategy, but this section has, I hope, made clear that it is worth having, notwithstanding cultural pluralism and our current lack of ritual infrastructure.

5.2.6 Prescribed Environments – Claim (11)

Given the effects that I have already discussed regarding attention, executive functions, emotions, and roles, it should be clear that one set of the avenues through which a prescribed ordering of our environments might help to support consistent moral behavior is by having effects on our moods and emotions, our capacity to direct our attention, and the salience of the morally relevant aspects of our self-concepts. In this section, I will present evidence that our environments can in fact have those kinds of effects, and so the Confucian strategy of designing our environments in certain ways.

One way in which our environments can effect us is through the quality of our social interactions. Hagop Sarkissian has recently argued that a response to the situationist challenge can come from Confucianism through its recognition that the quality of a social interaction can have behaviorally-relevant influences on people.89 In support of this idea, Sarkissian cites an experiment in which people rated their own emotional expressiveness, and in which high-scoring people were put into a room with two low-scoring people, which

---

found that the mood of the expressive individual influenced the mood of the less expressive individuals.\textsuperscript{90} Sarkissian notes that this finding is in line with work on moral contagion that shows that:

“1) in interpersonal contexts, people automatically and continuously mimic others, synchronizing their facial expressions, mannerisms, tone of voice, posture, etc.; (2) through a feedback mechanism, such mimicry elicits the relevant emotional states in the individuals at hand; and (3) emotions are thus transmitted and ‘caught’ by other individuals.”\textsuperscript{91}

By training individuals to attend to and control their emotions and their expression of emotions as well as prescribing certain emotional expressions during certain social interactions, the Confucian ritual strategy systematically places us in situations in which we can be influenced by others’ emotional states in ways that can, through the behavioral effects of moods and emotions, help to steer us toward morally correct behavior.

In addition to direct social interactions, we can also be influenced by our social situations by way of our broader environments, because of the way in which our environments can give us information about our social fellows, the norms they follow, and what they approve and disapprove of. In an experiment on littering behavior, it was shown that seeing a confederate litter made people less likely to litter a flyer than people in a control condition, but that the effect was far more pronounced in a clean environment than a dirty environment; the littering act activates a norm against littering, but doing so in a cleaner


environment signals a greater general compliance to that norm. A far stronger effect has been shown by the activation of injunctive norms rather than merely descriptive norms in a study in which the littering opportunity was presented to subjects who witnessed an experimental confederate actually picking up and throwing away a piece of litter; seeing somebody pick up a piece of litter caused a dramatic reduction in littering behavior regardless of the cleanliness of the environment. While a clean environment is a greater signal of the acceptance of an anti-littering norm among one’s social fellows than a dirty environment is, seeing somebody actively undo an act of littering more directly expresses a social norm against littering.

This all, I believe, speaks in favor of the Confucian strategy. By prescribing certain behaviors, certain forms of dress for certain occasions and certain arrangements of both private and public spaces in ways that are tied to behaviorally-relevant roles, Confucian rituals give us consistent ways of both making the moral more salient and signaling our acceptance of various norms. By wearing prescribed mourning attire or decorating our homes in a certain way, for instance, we can collectively signal to each other that we accept a set of norms for how to respect other people, and this signaling should help to support norm-consonant behavior through both the mechanism of social influence and the mechanism of increased norm salience. As I mentioned in 5.2.5.1, this sort of social signaling via rituals requires that we exist in a social environment in which people understand


the significance, which is not something that we currently have, as we are not in a society that has yet come to see the value of these kinds of rituals and put them into practice. That said, it is fairly easy to see how people could come to put them into practice within a circumscribed social circle, and then to encourage the practice to spread in ways that both make them more effective and allow more people to benefit from the increased social coordination.

Just as we can be affected by our environments because they give us social norm cues, we can be influenced by our environments because of ways in which they activate behaviorally relevant stereotypes or moods. It has been shown, for instance, that we can be primed in ways that affect our behavior through subconscious signals such as the presence of words relating to a certain feeling or behavior. The word-unscramble task that I discussed in section 5.2.5 has been used to show that we can be primed for either rude or polite behavior by showing that those whose word task was littered with words like “aggressive” and “rude” were then more likely to interrupt a subsequent conversation than those with a neutral prime, and those with words like “polite” and “patient” were less likely to interrupt than those with a neutral prime.94 In addition to words related to the quality of a particular behavior, it has also been shown that performing the same task with emotional terms can serve to prime certain behaviorally-relevant emotional responses. When participants in one study were given words like “hot” or “explode”, they expressed a significantly stronger anger response to a standardized provocation than those who were given a neutral prime, and those who were given words like “cool”, and “calm” showed a significantly weaker anger

---

94 Bargh, Chen, and Burrows, "Automaticity of Social Behavior".
response. It was also shown that the emotional control prime, while effective, “did not entail a cost in terms of negative emotion experience or maladaptive physiological responding”\textsuperscript{95}, meaning that it appears to be effective in a way that does not tax the same resources as conscious, explicit emotional control. In another study, a bucket filled with water and a citrus-scented cleaner induced people to more often choose cleaning as a future activity, and to leave fewer crumbs behind after being given a cookie to eat.\textsuperscript{96}

In addition to activating certain conceptually-mediated effects on our moods and thought processes, our environments can also more directly affect our moods in ways that are behaviorally relevant. An environment that induces a feeling of disgust, for instance, has been shown to increase the severity of moral judgments.\textsuperscript{97} Given our tendencies to feel motives of retribution and the way in which negative moral judgments of people can contribute to biased social though, it is in our moral interest to head off this effect, and so it may be in our moral interest to design our environments in ways that head off feelings of disgust. Our environments can also effect our behaviors by inducing other feelings. Pleasant smells, for instance, have been found to increase helping behavior, most likely because of their effect on mood and the link between more positive affect and increased helping behavior.\textsuperscript{98} Given these situational effects on behaviors and moods, it appears that the effort in arranging certain aspects of our environments that Confucian rituals prescribe


may also be an effective method of supporting moral behavior. We may, of course, need to learn a good deal more about what kinds of effects our environments can have, and this may lead us to design environments in ways that look very different from anything Confucius would have recognized, but the general strategy appears to be validated, and is something that we can put to use in a modern context with better and better effect as we learn more and more about the exact nature of these situational influences.

5.3 The Confucian Approach in Relation to Other Approaches

In this section, I will juxtapose the Confucian approach to moral cultivation with a few alternatives and make the case that the Confucian method has strengths that make it a valuable addition to our stock of methods of overcoming our given psychologies so as to allow us to behave in a more reliably moral manner. I will begin in section 5.3.1 by discussing Aristotle’s method of cultivating virtue, as laid out in the *Nicomachean Ethics*. I will then, in 5.3.2, discuss Mark Alfano’s recent proposal that we make virtue attributions of people as a means of inducing virtuous action and perhaps ultimately cultivating virtue in them. I will then move on to discuss, in 5.3.3, the prospect of biotechnological moral enhancement, something that has been strongly advocated by Julian Savulescu and Ingmar Persson.

In order to discuss how Confucianism measures up to each of these other proposed methods, I need to get clear on what criteria we should use when we perform that assessment. One relevant criterion is the actual effectiveness of the method. In judging the effectiveness, we need to look to the empirical evidence that is available to support the claim
that the method will have the desired effects, and that these effects are what we need in order to allow a person to behave in a more consistently moral manner. Another part of judging effectiveness is judging the completeness of the method. Perhaps the empirical literature gives us reason to believe that a method works as advertised, but the overall effect of implementing that strategy is still relatively limited compared to other methods. On this particular point, one thing that we should note at the outset is that, as I discussed in section 2.4, the more that we are capable of building cross-situationally robust dispositions, the less we need a mixed strategy that attempts to shape both individuals and environments.

However, adding elements of a mixed strategy, which the Analects does, as should be clear from Chapter 4, will tend to make a method more complete, regardless of exactly how possible or impossible it turns out to be to build a full complement of cross-situationally robust dispositions. This is because situation-engineering strategies will be necessary unless building robust, cross-situational dispositions is possible for everyone, and even in that case situation engineering strategies will still be effective, unless we discover both that building cross-situational consistency is the only real way to ensure compliance to one’s norms and that pursuing situation engineering strategies is somehow detrimental to that aim, which appears extremely unlikely given the situationist evidence.

A further criterion is what we can call the practicability of the method. In asking if a method is practicable we are asking whether it can easily and regularly be put into practice in our lives, whether we have sufficient opportunities and energy to put the method to work. Practicability is related to effectiveness, because if a method is not one that we have many opportunities to apply, it is like that it will be less effective than one that we can make a part
of our daily practical lives. Another criterion that we should consider, related to practicability, is the attractiveness of the method. If the method is overly burdensome, or if it involves practices that people would otherwise not want to perform, then a similarly effective method without that issue is clearly preferable and more likely to be widely utilized. On the flip side, if the method is itself an enjoyable activity, if it feels less like work than simply enjoying oneself, it speaks in favor of the method because pursuing it will feel like less of a trade-off. We are, after all, largely self-interested creatures with a strong near-term bias, especially if we are near the beginning of an attempt at moral cultivation, and any sugar that helps the moral medicine go down will allow the method to be more widely and regularly utilized.

Of these criteria, effectiveness is clearly the most important, because even if a method is practicable or attractive, we have no reason to take it up if we have no reason to think that it will actually do anything for us; it would just be one more way to spend our time. Effectiveness can, of course, be increased through completeness, practicability, and attractiveness, and can thus strengthen that particular reason for taking up one method rather than another, but attractiveness, for instance, can also stand alone as a reason to take up a method, especially insofar as people believe that we need not or ought not subordinate every aspect of our lives to the pursuit of moral perfection.

5.3.1 Vs. Aristotle’s Method of Cultivating Virtue

In the *Nicomachean Ethics*, Aristotle develops a theory of the virtues, including how to explain what makes something a virtue, why the virtues are good, how the virtues are
related to one another, and how we can come to exhibit the virtues. It is this latter aspect of his theory, the aspect that could be called his approach to moral cultivation, that I will concern myself with here.

Before I do, however, I need to clarify that Aristotle’s method is intended to cultivate virtues that are conceived of in a particularly robust, demanding way that goes beyond the lower bar of behavioral compliance that is my aim. For Aristotle, having full ethical virtue means having practical wisdom,99 so that one will not only have dispositions to do what one takes to be right, but also the ability to determine what it would be right to do. To commit a virtuous act, for Aristotle, means not simply performing an act that it would be right to perform, but to perform that act for the right reason100 and to make the choice to do so from a firm and unchangeable state.101 This is simply a higher bar than I am currently interested in, because it seems to me that fostering behavioral compliance to our norms is a difficult enough problem to solve without also placing these kinds of restrictions on how we can get there. So, rather than assessing how well Aristotle’s cultivation method fosters full Aristotelian virtue, I will assess it in its ability to cultivate something that is closer to what Aristotle calls continence, i.e. the ability to ultimately choose the right behavior even if that choice is the result of a sort of internal struggle in which one overcomes improperly aimed desires.102

100 NE 1144:25-31.
101 NE 1105a:30-34
102 NE 1150:34-35.
We should ask, though, whether this specific original intent makes the Aristotelean virtue cultivation project different enough from the Confucian project to make the juxtaposition somehow improper. I do not believe so, for the following reasons. First, despite the fact that I am primarily interested in the Confucian method as a way of cultivating behavioral compliance, there is a case to be made that the Confucian method was intended to clear a higher bar, rather similar to Aristotle’s, with behavior coming from a similar sort of firm and unchanging state\textsuperscript{103}, a similar sort of action only from the right reasons\textsuperscript{104}, and a similar sort of practical wisdom\textsuperscript{105}, so I am perhaps being overly cautious here in noting the differences. More to the point, though, whatever differences there are in terms of what, beyond successful behavioral compliance Aristotle’s and Confucius’ methods are intended to give us, both methods are at least still intended to be sufficient for getting our behavior to consistently comply with moral norms. As such, I see no reason why they cannot be differentially evaluated on their potential success for that.

While his theory of the virtues, the way in which they are grounded in our nature, and so forth is rich and robust, Aristotle pays less attention to answer the question of what method we are to use in order to cultivate those virtues. Daniel C. Russell believes that this is because Aristotle did not think that the “how” questions of morality were especially vexing or in need of answers that were not already available to his students and interlocutors as a result of asking and answering “how” questions in other contexts. As Russell explains:

\textsuperscript{103} Recall \textit{Analects} 2:4, where Confucius said he could give his heart and mind free reign without overstepping the line. There is a plausible reading of this passage that says that Confucius has reached a firm and unchangeable state.

\textsuperscript{104} Even if there would most likely be some significant disagreement between Aristotelians and Confucians over what those right reasons are.

\textsuperscript{105} There is a plausible understanding of “\textsl{yi}” in the \textit{Analects} as connoting a sense of what is appropriate to the situation.
“Aristotle does not offer any special theory of how we get better with respect to the virtues—and that, I think, is one of the most telling features of Aristotle’s approach: he offers no special theory of acquiring the virtues because he thinks that there is no special problem involved in understanding how virtues are acquired. On the contrary, Aristotle thinks of that process as a particular instance of something people do all the time: getting better at something through practice and training.”

To Aristotle, we acquire virtue qua persons in the same way we acquire virtue qua musicians or cooks, by practicing that particular craft so as to develop our skill in performing that craft:

“The virtues, on the other hand, we acquire by first engaging in the activities, as is also true in the case of the various crafts. For the things we cannot produce without learning to do so are the very ones we learn to produce by producing them—for example, we become housebuilders by building houses and lyre-players by playing the lyre. Similarly, then, we become just people by doing just actions, temperate people by doing temperate actions, and courageous people by doing courageous actions.”

When we play music in a certain way, we develop habits in how we play music. Similarly, to Aristotle, by behaving in certain ways, we develop habits in how we behave, and to have virtue is to have developed the right kinds of habits with respect to our behavior.

Developing these habits is not, to Aristotle, something that we need to work on the way we would need to, say, make time to play the lyre. We simply need to do the right things as we go about our lives, since “it is from acting as we do in our transactions with other human beings that some of us become just and others unjust, and from acting as we do in terrible circumstances and from becoming habituated to feel fear or confidence that some of us become courageous and others cowards.”

The way to cultivate a virtuous

---


107 NE 1103a:30-1103b1-2.


109 NE 1103b:13-17.
state, then, is to perform virtuous actions, since “states come about from activities that are similar to them.”\textsuperscript{110}

This habituation process is guided by what Aristotle takes to be our primary motivations, pleasure and pain.\textsuperscript{111} We need to learn to take pleasure in performing the right sorts of activities and to be pained at performing the wrong sorts of activities, so that we will do the right thing far more often than we do the wrong thing, and thus will build the habit of doing the right thing. This is one of the reasons why for Aristotle, having the right sort of upbringing is a necessary starting point for becoming virtuous.\textsuperscript{112} When we first start attaching pleasure and pain to certain activities, we need to be guided well so as to not be set up to make the wrong choices, which will instill the wrong kind of character, which will lead to further wrong choices, and so on.

I have four main criticisms of Aristotle’s method of cultivating virtue that I believe show the Confucian method to be a more compelling alternative by the criteria of practicability, completeness, and effectiveness. The first is that his approach does not countenance the way in which complex skills, like playing music or performing just actions, can be seen as applications of a number of different capacities that we are capable of isolating and finding efficient methods to practice. To perform a generous action, for instance, one must notice that one has an opportunity to be generous, feel at least motivated enough to give to consider it as an action possibility, and possibly override any strong motivations to keep something for oneself that would prevent one from giving it away. For

\textsuperscript{110} NE 1103b: 20-21.
\textsuperscript{111} NE 1104b:26-28.
\textsuperscript{112} NE 1104b:11-13.
every complex skill like this that one can try to hone, there are a variety of pedagogies that can be developed for instilling that skill, and the difference between pedagogies often amounts to a difference in the ordering and training methods for instilling more basic skills.

For instance, in teaching percussion, I often aim to instill sub-skills that will make my students better drummers, like concentration and what I call embodied visualization, and the exercises that I give them to instill those sub-skills often involve no actual drumming whatsoever, but get results none-the-less. This sub-skill approach makes sense in light of the way in which, as I discussed in section 5.2.1, certain skills can become more and more automatic, freeing up attentional resources for focus on more and more layers of the skilled performance. So, the sub-skill approach is not only a finer instrument for correcting deficits in our macro-skills, but it is also a natural pedagogical approach considering the role of automaticity in skills. It is also the case that, as regards moral skills, learning certain sub-skills like putting others at ease, acting so as not to incur resentment, etc. can give one a different social playing field that may make certain other skills easier to perform, or may even open up the possibility to use higher-level morally-relevant skills, like rallying people around a cause.¹¹³

In the Confucian approach, people work to train certain fundamental capacities that influence our ability to do just, generous, or helpful actions. Making a sacrifice at every meal is not in and of itself a just action, because it is not an instance of dealing with another person, fairly or otherwise, but it does train a more basic ability and tendency to detach from what one has, to be okay with having less than one otherwise could have had, and the ability

¹¹³ I am grateful to David Wong for helping me to articulate this benefit of the sub-skill approach.
to do this can contribute to a person acting justly in justice-relevant situations. The benefit of this kind of approach, of working at the level of the attentional, executive, and emotional roots of behavior rather than at the level of behaviors themselves, is something that is borne out through much of the empirical literature that I cited in section 5.2; training our attentional capacities and executive functions allows for more consistently moral behavior, but the practices that train those capacities are not, directly, instances of choosing virtuous behaviors over vicious ones.

The second criticism is that if the method involves simply waiting for opportunities to arise to exercise our skill in choosing behaviors that correspond to particular virtues, then our path to virtue will both be slower than it could be, and it will be more subject to luck than a method that gives us more opportunities for practice. Opportunities to perform various virtuous acts are not evenly and regularly distributed to people. This means that on the Aristotelian approach, the virtue that one can accumulate depends in large part on the opportunities to practice virtue with which one is presented. Somebody tempted more often to injustice will have more opportunities to be just; somebody tempted more often with food or wine or sex will have more opportunities to exercise temperance. Rituals, on the other hand, give us a way of structuring our habit-building so that it can happen in a regular, consistent manner. If the case that I made in section 5.2 holds, then people need not be given an opportunity to be unjust to improve and choosing just actions; one can, instead, structure one’s day so that one is presented with a number of opportunities to perform rituals that develop morally-relevant skills that make one more capable of just action. So, the
Confucian method on this point is more practicable, and, as a result, should be more reliably effective.

The third criticism is that the picture of our behavior as resulting largely from our taking pleasure in certain actions and being pained at performing others fails to account for a number of other factors that we now know to have influences on our behavior. As I have discussed in Chapter 2 and the present chapter, our behavior is affected not only by our preferences and conscious motivations like pleasure and pain, but also by situational factors that affect us in ways that we do not consciously register and self-concepts that often induce us to do things in which we take no pleasure and even find painful because to behave so is to behave as who we believe we are. In not taking situational factors into account, marshalling them as a part of the cultivation method, or attempting to equip people with an awareness of situations and their effects on people’s behaviors, the Aristotelian method loses out on opportunities to affect behavioral outcomes, and gives us little by way of a method of overcoming situational influences.

The fourth criticism that I have is that Aristotle’s method is not going to be much help to people who did not have the right kind of upbringing. If one was not taught proper habits, then an approach that amounts to continuing to habituate the good habits that one has will not be useful. The Confucian approach, on the other hand, gives even those among us without the correct pasts a set of tools for understanding and reshaping the mental habits that underlie our behavioral habits. Much of the work on the benefits of mindfulness practices that I have cited, for instance, has been performed in therapeutic contexts, helping people with behavioral difficulties, many the result of childhood trauma or other improper
upbringings, to reform their behavioral patterns through greater attention, emotional insight, executive control, and the different emotional and mood states that come along with them. The people who most need methods of moral cultivation are often those who did not receive the necessary first step in Aristotle’s process. In giving us something that is practicable and effective for these people, the Confucian approach gives us something that can be more productive in pushing people, generally, toward consistently correct behaviors.

5.3.2 Vs. Factitious Virtue

In a recent book, Mark Alfano has presented a new proposal for a method of cultivating moral compliance. This proposal is part of an attempt to defend virtue ethics from the situationist challenge that attempts to undermine virtue ethics by showing that people do not have character traits. His response to this challenge is to argue that we still have good reasons to make virtue attributions “even in the absence of what would ordinarily be called sufficient evidence”\(^\text{114}\). We have reason to make these attributions because the empirical literature, according to Alfano, can be used to justify the claim that “plausible, public attribution of virtuous traits induces in the target of the attribution both the identification with those traits and a belief that others expect him to behave in trait-consonant ways, which in turn leads to trait-consonant conduct.”\(^\text{115}\) This induced trait-consonant behavior is what Alfano calls “factitious virtue”\(^\text{116}\). The possibility of inducing


\(^{115}\) Ibid., p. 82-83.

\(^{116}\) Ibid., p. 10.
factitious virtue through a virtue attribution makes that virtue attribution more worthy both by making it more true and by giving the attributive act a significant pragmatic value.

In addition to defending virtue ethics from the situationist challenge, Alfano’s project is meant to highlight what he calls “moral technology”117 and what I have been calling cultivation methods, i.e. the tools and techniques that we can use to shape our psychologies so as to enable us to behave in a more consistently moral manner. He advances factitious virtue not only as a justification of virtue judgments, but also as an entry in our stock of cultivation methods: “Aristotle thought people became courageous (or near enough) by acting courageously; I contend that people become courageous (or near enough) by being called courageous”.118 While factitious virtue is different from full virtue, Alfano does think that “a case of factitious virtue may, in the long term, bleed into outright virtue”119, so long as the virtue is understood in “behavioral, motivational, and cognitive terms”120, leading the target of the attribution to associate the attribution not merely with behavioral compliance, but also with having the appropriate sensitivities to a virtue’s characteristic reasons and having the right motives. As I have mentioned, I am only interested here with assessing these methods’ capacities to ensure behavioral compliance, so what Alfano calls “outright virtue” seems to be more demanding than what I am looking for a method to fulfill. However, insofar as our desires and our sensitivity to reasons affect our behavioral outputs, his favored, more demanding conceptions of virtue terms may enable his method to better clear that lower, behavior-only bar by ensuring greater behavioral

117 Ibid. p. 6.
118 Ibid., p. 10.
119 Ibid., p. 94.
120 Ibid.
compliance as opposed to merely adding a layer of “right reasons” frosting to the good behavior cake.

The major claim I want to make about the factious virtue approach vis-à-vis the Confucian approach is that what speaks in favor of the factitious virtue approach speaks in favor the Confucian approach, but that the opposite is not true. While virtue attributions and role attributions are both effective, I want to contend that the Confucian approach is far more complete and somewhat more practicable. The evidence that speaks most strongly in favor of the belief that the factitious virtue approach is effective comes from labeling theory in psychology, which I argued in section 5.2 also speaks in favor of the Confucian approach because of its reliance on role attributions of a similarly public, plausible manner. The evidence bears out the effectiveness of role attributions just as clearly as it bears out the effectiveness of trait attributions. This is something that we should expect considering our best explanation of why labeling is as effective as it is, namely that labeling influences our self-concepts, which in turn affect our behavior in various ways. Our social role attributions, like trait attributions, are bound up with our self-concepts and tell us how others expect us to act, what we are expected to want, and what reasons to which we are to be sensitive.

I think that social role attributions have a few virtues that trait attributions do not. First, attributions have to be plausible to be effective, and social role attributions have to clear a much lower bar to be plausible; generally, they are completely plausible because social role attributions are thick terms with normative and factual content, and the factual content is typically immediately obvious like “I have parents” or “I am a teacher”. Role attributions are also made more plausible by the fact that, as I discussed in section 5.2.5, role terms are
more heavily weighted toward expressing normative expectations versus expressing normative success than trait terms are. This means that they are more likely to be believable in cases in which somebody is falling short. If, for instance, I am not particularly caring caretaker, being attributed the trait “caring” may be too implausible to affect my self concept, but being reminded continually that I am a caretaker may be more believable because it plays more of a role of expressing normative expectation, giving it a behavior-affecting foothold in my self-concept that “caring” might not achieve. This easier plausibility, combined with the effectiveness of the normative expectation aspect of role terms that I also discussed in section 5.2.5, means that there are far more situations in which an effective role attribution can be made than an effective virtue attribution, which likely means that progress on and maintenance of our self concepts and our resultant behaviors should be more consistent.

Second, social role attributions are also already very deeply bound up with our self-concepts. This means that the strategy for imbuing our self-concepts with the right sort of normative alignment is importantly different in the Confucian approach than in the factitious virtue approach. One takes terms heavy in normative content and attempts to make them a part of our self concept, and the other takes terms that largely inform our self-concepts and attempts to engineer them to have effective normative content. Which of these strategies is more effective is an empirical question that still needs to be further studied. It does seem to me likely, though, that changing how we are affected by a term with which we already identify would present fewer hurdles than incorporating a new term into our self-concept. This seems to me to be especially the case when people’s self-concepts are
too negative for virtue attributions to be plausible; in such a situation, I do not see how the virtue attributions can gain a foothold into our self-concepts. Role terms, on the other hand, are already a part of our self-concepts, and they are also already carriers of normative content that are used to guide our social behavior, so the only real work to be done is to make attributions of those terms hit home a little differently.

Beyond the relative effectiveness of role and trait attributions, one of the advantages that the Confucian approach has over the factitious virtue approach is that the rituals give us a structured opportunity to make the role attributions, which makes it more practicable than the factitious virtue theory in a few significant ways. First, a system of rituals that situate us in our roles allows us to regularly and reliably update our self-concepts by way of role attributions. Instead of simply going through our lives until happenstance leads us into situations in which a public, plausible trait attribution can be made, we can structure our social interactions in ways that remind ourselves throughout the day that we are siblings, friends, colleagues, teachers, citizens, partners, teammates, etc.

Second, rituals give us an opportunity to maintain our role concepts. By performing certain behaviors in certain ways with certain emotions and attending to certain details while consciously embodying our roles, we associate the role terms with the right sorts of dispositions. By doing this regularly, we do what we can to protect the terms from their natural evolution so that our attributions continue to have the right effects. Factitious virtue alone gives us no delivery or maintenance system for our attributions. We could, perhaps, give ourselves reminders of our moral education and make sure that we still think of the terms in the same way, but rituals can integrate these reminders into our lives in more
automatic ways, because performing the rituals helps to constitute the content of the role terms for us, making it easier and, for what it is worth, more enjoyable, to perform the necessary maintenance of the terms. Given the discussion in section 5.2.5.1, the success of rituals in being able to perform this maintenance, of course, will be relative to however large of a social circle that we can get involved with that project, so the less cultural fragmentation we have, the more widely we can maintain these concepts. But, considering that cultural fragmentation also affects trait attributions, factitious virtue and the Confucian approach both have to contend with this difficulty. I believe that this helps to highlight, however, one of the further benefits of the Confucian approach, namely that rituals themselves can be attractive in ways that may allow them to spread and become more homogenous over larger and larger social circles, and this provides a further mechanism, beyond discussion of we ought to be using our terms, to help foster a beneficial consistency.

Furthermore, because rituals are social in nature and define roles for participants collectively, the system of rituals is more likely to keep people on the same page with respect to the content of the role terms. Because the enactment of a ritual will so often situate multiple individuals in social roles, the enactment of a ritual gives more per-capita concept maintenance than a trait attribution does. This effect, of course, as per the discussion in section 5.2.5.1, will be relative to whatever social consensus we can garner regarding the social roles and rituals, like how in the case of factitious virtue, the effect will be relative to whatever social consensus we can gain regarding trait terms. Even noting this shared weakness, though, we can see that the Confucian approach has a further advantage, namely the greater stability of a structure of inter-defined social role terms compared to the stability
of trait terms. An on-target attribution of “courage” helps to maintain that term; an on-target attribution of “teacher” helps to maintain “teacher” and “student”.

Aside from these points, the Confucian method is also a good deal more complete than the factitious virtue approach, simply because the training works directly on some of the basic capacities that constitute our agency, like our attention and our executive function. As I have shown, strengthening these has a significant impact on our ability to behave in the pro-social ways in which all plausible moralities tell us we ought to act. As such, we should prefer a moral cultivation strategy that involves both the deliberate cognitive and affective training and the various effects that come from morally-laden labeling.

The Confucian approach gives us something that factitious virtue does not, because it situates labeling in the context of a system that both makes the labeling more regular and effective, and works to give us an array of skills that should help us to more consciously live up to the labels with which we identity. Perhaps factitious virtue could be incorporated into this kind of large system, with rituals allowing for and standardizing virtue attributions, or perhaps the insights could be used to help bolster a program of Confucian mindfulness training by labeling people as “diligent” when they practice or “mindful” when they appear to be in the right sort of state. Either way, the result would contain a significant Confucian contribution.

5.3.3 Vs. Biotechnological Moral Enhancement

Another possible approach for moving people in a pro-moral direction that has been discussed in recent work is biotechnological moral enhancement. The lead proponents of
this sort of approach have been Ingmar Perrson and Julian Savelescu. In their book, *Unfit for the Future*, they argue that “human beings are not by nature equipped with a moral psychology that allows them to cope with the moral problems”\(^\text{121}\) that arise from our modern conditions of life, wherein technology has enabled us to endanger our very existence. Because of this, they argue, “the current predicament of humankind is so serious that it is imperative that scientific research explore every possibility of developing effective means of moral bioenhancement, as a complement to traditional means.”\(^\text{122}\) They acknowledge that we “could become more morally motivated, morally enhanced, by a more thoroughgoing application of the traditional methods of moral education”\(^\text{123}\), but that “these methods appear to have had modest success during the last couple of millennia”\(^\text{124}\), giving us reason to explore “supplementary techniques of moral enhancement, such as pharmaceutical drugs or genetic modifications.”\(^\text{125}\)

Their case ultimately rests on the fact that there is research showing that certain biomedical interventions have particular pro-moral effects. One such intervention is the elevation of levels of the hormone oxytocin via a nasal spray. Persson and Savelescu cite a study in which participants played a cooperation game that involved making decisions that were used to operationalize both the showing of trust and the showing of gratitude. In this


\(^{122}\) Ibid., p. 2.

\(^{123}\) Ibid., p. 9.

\(^{124}\) Ibid., p. 9.

\(^{125}\) Ibid., p. 9.
study, those “administered oxytocin exhibited significantly more trusting behavior.” This result is bolstered by another study showing that “receipt of a signal of trust by the trustee is associated with a spike in oxytocin levels and that the degree of trustworthiness exhibited by the trustee is positively and significantly correlated with the oxytocin level.” They admit, however, that other studies show oxytocin to have prosocial effects by apparently strengthening our sense of in-group bonds, which also strengthens outgroup animus or indifference.

Another apparently effective intervention that they cite is the use of Selective Serotonin Reuptake Inhibitors (SSRIs) to increase the amount of the neurotransmitter serotonin found in a person’s synapses in various areas of the brain. In one study that Persson and Savelescu cite, among people who were playing a dictator game, in which a person is given money and must unilaterally decide how much of it to give to another participant, those who were given an SSRI split the money more fairly than controls. In another study that they cite, it was shown that “depletion of precursor of serotonin (tryptophan)… leads to lower rates of cooperation in the prisoner’s dilemma game,” suggesting that SSRI’s, by increasing serotonin levels, can increase people’s likelihood of

---


exhibiting cooperative, rather than defective, behavior. They also cite the effectiveness of certain drugs like Ritalin and Adderall in allowing people to have greater impulse control than they otherwise would, something that can help in “reducing violent and anti-social behavior.”

The first critical point that I would like to note about this approach is something that Persson and Savelescu have already acknowledged. While they cite a fair number of studies that show that certain individual chemical interventions are capable of affecting our decision-making processes in ways that push us toward more pro-social behavior, and while this should lead us to believe that moral bioenhancement represents a genuine possibility for the future and an opportunity for interesting and effective research, they still admit that “moral bioenhancement worthy of the name is practically impossible at present,” largely because the relevant sciences are still in their infancy. So, whatever the eventual prospects for moral bioenhancement in the future, it appears that we will still need to rely on the best of whatever traditional methods of moral enhancement we have available.

What I would add to their admission is that a future in which we can morally enhance ourselves biomedically may be far enough out that, given appropriate focus, we will figure out ways of applying more traditional methods of moral cultivation that allow us to overcome both our mundane impediments to virtue and a potential global moral catastrophe before biomedical moral enhancement becomes a legitimate possibility. Part of the reason why biomedical moral enhancement is at this point a distant dream is that the actual neurochemical processes underlying human behavior, human moods, and so forth are

---

incredibly complicated. As the research on the effects of oxytocin and SSRI’s shows, altering our brain chemistry in one particular way can have a number of effects, some of which are pro-moral, and some of which are actually anti-moral. Between the number of variables and the number of different effects that changes in those variables can produce, teasing out the right sorts of interventions may be an intractable problem.

As of right now, the development of more traditional methods of altering our psychologies in pro-moral directions is outpacing not only the development of biomedical enhancement, but also our ability to understand why the traditional methods work as well as they do. While mindfulness practices, for instance, have yielded a number of positive results in therapeutic settings, it is also widely accepted among psychologists who study it that “the processes underlying these clinical effects are presently not well understood.”

This makes sense considering the fact that when we are able to demonstrate that there is an effect at the behavioral level, i.e. that people who are called tidy tend to act more tidily, the descriptions of these events at lower and lower levels of description become more and more complicated. Since being able to explain something at those lower levels of description is a necessary part of designing interventions at those levels, we should expect that the development of behavior-based investigations will continue to outpace the development of interventions at those lower levels at those lower levels.

A related worry is that, in focusing so heavily on finding interventions that we can perform on individual people and the biochemical processes that occur within individual people, Persson and Savelescu may be losing sight of much of what is actually relevant in

finding effective methods of moral enhancement. This is because the already complex neurochemical processes that they are discussing are always occurring in the context of yet more complex social systems and processes, in ways that not only help to determine what socio-behavioral effects different biotechnological interventions may have, but that also themselves have profound effects on the very neurochemical processes that Persson and Savelescu are proposing that we alter. In other words, the variables that they seek to change both influence and depend on variables that are thus just as relevant, but which their approach seems to simply ignore. These added variables, of course, vastly compound the difficulty of determining how the kinds of chemical tweaks that they are proposing might yield the desired moral enhancement results. For these reasons, finding an effective method of moral enhancement/cultivation that focuses so heavily on the neurochemistry of the individual seems less likely to be successful than one that, like the Confucian approach, countenances the various effects of situating an individual in broader, complex social systems of different sorts, and focuses on interventions at that level.

I also think that we should give traditional methods far more credit than Persson and Savelescu seem to be giving them when they say that these methods have only had modest success over the last two thousand years, for three reasons. The first is that the moral success of humankind has actually been pretty remarkable. In a recent book, Steven Pinker has made the case that the decline of violence as an occurrence in human life on scales both large and small “is an unmistakable development, visible on scales from millennia to years,
from the waging of wars to the spanking of children.”  

This moral trend is, of course, the result of a variety of factors from changes in political culture toward democracy (“All of the first complex civilizations were absolutist theocracies which punished victimless crimes with torture and mutilation.”), to greater prosperity, better nutrition, and the greater intelligence and collective prudential reasoning that prosperity leads to, to more recent “efforts to stigmatize, and in many cases criminalize, temptations to violence [that] have been advanced in a cascade of campaigns for ‘rights’”.  

Reading the descriptions of the routine violence that permeated people’s lives even a couple of hundred years ago, it is undeniable that we have made massive moral progress through traditional moral education and cultivation, moral philosophy, social and political development, and other means that did not require biotechnological enhancements.

The second reason we should not be so pessimistic about traditional methods of moral cultivation is that we have only recently gotten to the point where we can get together globally as a species and compare notes about all of the various moral education programs, cultivation methods, and moral technologies that people have employed during that period in which we have collectively acted more and more in accordance with the moral norms on which we mostly seem to agree. The project of putting those moral methods to the test in scientific inquiry and seeing which ones are likely to be more effective, explaining how they are effective, finding what effective methods have in common, and redesigning more and more effective methods on the basis of what we have figured out has only recently begun. It

135 Ibid., Kindle Locations 3082-3083.
136 Ibid., Kindle Locations 8468-8469.
may be the case that traditional methods, properly studied and pursued, can help us to continue our species’ upward moral trend, pushing us past our lack of concern for distant others and our tendency to tragically ruin the commons in the same way they have pushed us past slavery, dueling, and public, torturous executions.

Beyond these points about practicability in the here and now and the effectiveness of traditional methods of moral cultivation, I think that we will continue to have reasons to prefer the Confucian approach, or at least to think that it should still be a valuable resource even as our ability to control our behavior through scientific interventions becomes a possibility. Even if a technology like genetic engineering becomes widely available and effective, it will still only work to tip the odds for any given individual in favor of moral behavior. If people are, say, genetically or biochemically modified to have certain emotional responses a little more or a little less strongly, or to have greater executive control, and this makes their chances of exhibiting good behavior greater, we still need to provide the right kind of environment to instill norms, maintain the right sorts of self-concepts, train and maintain emotional skills, and so on, in order for them to actually become people who behave in accordance with our shared moral norms. They still need to learn what to pay attention to in their environments, understand the (modified) emotional repertoire that they have, and understand how to go with feelings conducive to moral behavior and let go of feelings conducive to immoral behavior. In other words, just as genetically engineered, chemically-enhanced athletes still need methods of training for their sports, genetically engineered, chemically-enhanced moral agents will still need the kind of training that the Confucian approach can give us.
References


Ayduk, Ozlem, Rodolfo Mendoza-Denton, Walter Mischel, Geraldine Downey, Philip K. Peake, and Monica L. Rodriguez. “Regulating the Interpersonal Self: Strategic Self-


Makris, Nikos, et al. “Cortical Thinning of the Attention and Executive Function Networks in Adults with Attention-deficit/Hyperactivity Disorder.” *Cerebral Cortex* 17, no. 6 (2007): 1364-1375.


Watson, John. The Philosophy of Kant as Contained in Extracts from His Own Writings. New York:: MacMillan and Co., 1901.


Biography

Daniel J. Stephens was born in Grand Haven, Michigan on August 11th, 1983. He grew up in Grand Rapids, Michigan in circumstances that quickly taught him that most everything happening both around him and within his mind was in dire need of questioning and ameliorative revision. In his sophomore year at Grand Valley State University, he found academic philosophy and decided that it was an appropriate professional vehicle for the kind of investigations that he wanted to pursue. He received a B.A. in Philosophy from Grand Valley State University in 2006, and an M.Phil in Philosophy from the University of Hong Kong in 2008. While he was a student at the University of Hong Kong, he was the winner of the Graduate Student Paper Award at the 2007 annual conference of the Society for Asian and Comparative Philosophy. For one year, beginning in fall of 2008, he was Adjunct Professor of Philosophy at Grand Valley State University, in which time published an article, “Confucianism, Pragmatism, and Socially Beneficial Philosophy” in the March 2009 edition of the Journal of Chinese Philosophy. In his time studying at Duke, he also devoted a portion of his time as a percussion instructor and mentor for the Legends Performing Arts Association, a non-profit youth arts-education organization based in Kalamazoo, Michigan. He received a Summer Research Fellowship from Duke University in both 2013 and 2014. Upon leaving Duke, he has been awarded a postdoctoral fellowship at the Center for East Asian and Comparative Philosophy at the City University of Hong Kong.