PROJECTIVISM PSYCHOLOGIZED: THE PHILOSOPHY AND PSYCHOLOGY OF DISGUST

By

DANIEL RYAN KELLY

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Philosophy

written under the direction of

Stephen P. Stich

and approved by

Stephen P. Stich

Alvin I. Goldman

Brian McLaughlin

Colin Jager

New Brunswick, New Jersey

October, 2007

ABSTRACT OF THE DISSERTATION

Projectivism Psychologized: The Philosophy and Psychology of Disgust

by DANIEL RYAN KELLY

Dissertation Director:

Stephen P. Stich

This dissertation explores issues in the philosophy of psychology and metaphysics through the lens of the emotion of disgust, and its corresponding property, disgustingness.

The first chapter organizes an extremely large body of data about disgust, imposes two constraints any theory must meet, and offers a cognitive model of the mechanisms underlying the emotion. The second chapter explores the evolution of disgust, and argues for the Entanglement thesis: this uniquely human emotion was formed when two formerly distinct mechanisms, one dedicated to monitoring food intake and protecting against poisons, the other dedicated to protecting against parasitic infection, where driven together until they became functionally integrated. The third chapter explores the sorts of acquisition mechanisms that could account for the patterns of individual and cultural level variation we find with disgust elicitors. It argues for the Empathic Acquisition thesis, which holds that one important route for the social acquisition and transmission of disgust elicitors is linked to empathic recognition of facial expressions of the emotion. The fourth chapter builds on the Entanglement thesis, and embeds the emotion of disgust in gene-culture coevolutionary theory and the tribal instincts hypothesis. The Co-opt

thesis is defended, which maintains that disgust was co-opted to play an important role in our moral psychology, particularly in our cognition of social norms and ethnic boundary markers. In doing so, however, it brings to bear many features initially linked to poisons and parasites. This explains the puzzling and troublesome character of moral judgments linked to disgust.

After shifting gears from psychology to metaphysics, the fifth chapter recasts the Humean tradition of projectivism in the terminology of cognitive science. Using examples such as disgust, I argue that a psychologized projectivism is able to make sense of the idea that some properties are projected onto the world, rather than found there to begin with. The final chapter criticizes three other accounts of the property of disgustingness, two inspired by functionalism in the philosophy of color, one inspired by fittingness accounts in metaethics. I argue that none provide nearly as satisfactory an account of the property as the psychologized projectivism articulated previously.

TABLE OF CONTENTS

LIST OF FIGURES

Introduction

Pick a random person off the street and ask him to name five disgusting things off the top of his head, and you are likely to get an earful about filth, disease, death, bugs, and perhaps the mention of some sort of exotic food he finds particularly unpleasant, like raw fish or lima beans. These are the types of things most commonly associated with disgust, and they are concrete and almost brutely physical. The experience of the emotion itself, as opposed to the things that commonly induce it, is also deeply primal: the visceral sense of revulsion, the slight feeling of nausea in the gut, the worries about physical contact and contamination, the gaping facial expression that could so easily tip into actual retching. The response and the sort of things that typically induce it appear to be matched in their involvement with the body, the organic and physical, and the concrete.

On the other hand, consider Henry Higgins' startlingly strong reaction to Eliza Doolittle's diction when they first meet in the play *Pygmalion*:

> "A woman who utters such depressing and disgusting sounds has no right to be anywhere—no right to live. Remember that you are a human being with a soul and the divine gift of articulate speech: that your native language is the language of Shakespear and Milton and The Bible; and don't sit there crooning like a bilious pigeon."

Professor Higgins appears to have been wholly disgusted by nothing more than what he takes to be an improper accent. Most of us do not have such refined sensibilities that we are wont to be revolted by mere pronunciation, but the fact that something as abstract as idiolect could also induce disgust is telling, if not totally unfamiliar. In addition to its focus on the slime and filth of the physical world, disgust also rises above the mean things of the earth and involves itself in more abstract matters as well. In reporting that

the bourgeois thought that "the lower classes smell", George Orwell was arguing that for all the highfalutin debate and reasoning about political theory, one of the most difficult hurdles to achieving real social equality is that the bourgeois are secretly disgusted by the working classes. Appeal to smell can often be avoided when it comes to political opponents or rival groups – the very ideology and value system of those with whom we are set against can come to be deeply disgusting. In such cases, disgust can even take on a moral valence.

The arena of disgusting things ranges from the concrete and physical to the abstract and social, but it exhibits diversity in other ways as well. We are all disgusted by something or other. Common sense and casual observation suggest that there is great variance in what people find disgusting. Different things will disgust people with different sensibilities and different cultural backgrounds. Often each of us has our own personalized and idiosyncratic objects of disgust as well. One man's treasure is another man's trash.

0.1 Distinguishing Three Projects

Our random person on the street, indeed a random person found on a university campus, is likely to be surprised, or even skeptical, that the disgust could be of much theoretical interest, or the object of serious scholarship. This is understandable, perhaps, but a swell of recent work has raised the emotion from relative obscurity to new levels of visibility, and novelty alone does not account for the attention. The emotion of disgust has moved to a place of central importance to issues that lie at the intersection of philosophy and psychology. Even at this crossroads of overlapping concerns, though, we can distinguish three distinct projects. Each of these have in common that they touch on

the emotion, but each differs with respect to why disgust is relevant to its aims. Indeed, each project can thus be identified by its central goals and methods.

We might call one project associated with disgust the *Normative Project*. Those engaged in the Normative Project are likely to be addressing a cluster of issues that center on the question of whether disgust should enter into various decisions and evaluations, and if so, how it should be dealt with. These include normative questions concerning how feelings of disgust should be weighted in our moral deliberations. If we are attempting to achieve a state of ideal reflective equilibrium, for instance, what should we do with the fact that we are disgusted a particular social practice? Moreover, how should our legal system and other institutions deal with feelings of disgust? If a substantial majority of the population is disgusted by a social practice, how should this fact impact on the operation of our social institutions?

Some of the specific issues that come up in these types of debates include how feelings of disgust should influence the determination of culpability (can extreme feelings of disgust mitigate the responsibility one has for one's actions, in the way extreme feelings of rage do in cases of temporary insanity?) and meting out of punishment (if someone commits a crime that is not just illegal but repulsive, should this be reflected in a more stringent sentence?) In the past, disgust has been appealed to in the identification and legal definition of obscenity. More recently, disgust has played a role in ethical arguments about abortion, stem cell research, human cloning and homosexuality, gay rights and same sex marriages.

On the one hand, some of those engaged in the Normative Project take disgust to provide us with valuable information about the "naturalness" or "unnaturalness" of

certain practices. Accordingly, they suggest that we should let that information guide our assessment of those practices. For instance, the social historian and legal scholar William Miller (1997) entertains a view like this in his book *The Anatomy of Disgust*, which has had a wide-ranging impact throughout the humanities. The bioethicist Leon Kass (2002) has enjoyed a different kind of influence as former chair the President's Council on Bioethics for President George W. Bush. There, as a well as in his book *Life, liberty, and the defense of dignity: The challenge to bioethics,* Kass maintains that there is a certain "wisdom" in repugnance, elaborating that "in crucial cases...repugnance is the emotional expression of deep wisdom, beyond reason's power fully to articulate it." He even goes so far as to say "shallow are the souls that have forgotten how to shudder".

On the other hand, there are those who see feelings of disgust as merely expressing our own anxieties, and stemming from a deep seated but unreasonable repugnance of the organic, mortal body. These researchers argue we should discount those feelings in our deliberations about both morality and the law. Most prominent amongst those to defend such a position is Martha Nussbaum (2004), in her book *Hiding From Humanity: Disgust, Shame, and the Law*. She nicely summarizes her view in *The Chronicle of Higher Education* (August 6, 2004), where she muses:

> "Does disgust, then, contain a wisdom that steers law in the right direction? Surely the moral progress of society can be measured by the degree to which it separates disgust from danger and indignation, basing laws and social rules on substantive harm, rather than on the symbolic relationship an object bears to our anxieties."

Interesting and timely as these questions are, much of what I will have to say in this dissertation will speak to issues in the Normative Project only indirectly. Many of the normative arguments presuppose views on what disgust is, what it does, and how and

why it does it. My main concern lies with questions such as these, which fall in the domain of the other two projects.

We might call the first of these the *Metaphysical Project*. Those engaged in this project are often interested in disgust, or more often the property of disgustingness, because it can serve as a model to shed light on some of the core issues that arise in debates in metaphysics. These include issues about ontology, the status of various properties and how to locate them in nature, and what it might mean to say they are response dependent, dispositional, or projected. Metaethicists in particular have been paying attention to disgust and disgustingness, and using it to illuminate the structure and semantics of evaluative discourse (or content of evaluative experience) more generally. When people argue about whether or not something is disgusting, what are they arguing about? Are they talking past each other? How might we determine who is right? What are the truth conditions of ascriptions of disgustingness?

For instance, D'Arms and Jacobson (2000, 2005) agree with Wiggins (1987) and Blackburn (1994) that discourse about disgustingness shares properties that any account of moral discourse must be able to make sense of. That is, they agree that claims about disgustingness, about whether or not something is disgusting, are analogous to claims about morality, about whether or not some action is morally right, or good, or virtuous: both types of claims are thought to be both interpersonally authoritative and essentially contestable. Metaethicists advance these as two very general features that any viable account the semantics of moral discourse should accommodate.

Roughly speaking, interpersonal authoritativeness is meant to capture the idea that an ascription of disgustingness to some entity does not behave as if it were merely a

report of individual dispositions to be disgusted by the entity. Rather, a statement like "that Whopper is disgusting" also implicitly carries with it the implicit claim that others ought to be disgusted by it as well, and that if they are not disgusted by it, they are missing something. Essentially contestable, on the other hand, is meant to capture the idea that whether or not an ascription of disgustingness to some entity is correct resists being definitively settled one way or the other. Rather, such claims can be subject to criticism and debate. Indeed, according to Wiggins, the very function of essentially contestable concepts requires their application remain open to a particular kind of normative influence: the giving and taking or reasons in favor of each.

Similarly, disgust and disgustingness have been used, again as something of a simplified model, to illuminate issues about the ontology and metaphysics of certain types of response dependent properties more generally. In particular, those seeking to extend the Humean sentimentalist tradition have taken an interest in the emotion. The basic idea of sentimentalism is that evaluative concepts and properties, including moral ones, crucially depend (somehow) on the human sentiments. Many of the main themes of the sentimentalist tradition are often traced back to David Hume, who thought our moral and aesthetic judgments are grounded in our feelings of approbation and disapprobation.

More recently, John McDowell (1985, 1987) found a small and rare bit of common ground with J.L. Mackie (1977) in agreeing that disgustingness is a paradigm example of a property that is projected onto the world by the mind. This idea of projecting, which we will explore at greater length in the later chapters of the dissertation, is meant convey that in such cases, what we unreflectively take to be features of the external world are in reality projected onto the world by our minds, similar to the way a

film projector adds colored images to a screen that is otherwise blankly white. In his 1987 paper, McDowell remarks that it would be a "confused notion" to think that "disgustingness is a property some things have intrinsically or absolutely, independently of their relations to us". He maintains this is the case despite the fact that the phenomenology associated with disgust "presents itself as a matter of sensitivity to aspects of the world". If any property is worthy of a projectivist treatment, both seem to agree, disgustingness is.

In relying so heavily on the emotions in her metaethical accounts of morality, it would clearly behoove a sentimentalist to know as much about the nature of the emotions as she can. As D'Arms and Jacobson (2000) put it, sentimentalists: "need to look at each particular emotion more closely in order to determine the nature of its internal structure", for "it is necessary to examine our actual emotions piecemeal, in order to articulate differences in how each emotion presents some feature of the world to us when we are in its grip."

The relevance of this type of information about the emotions quite naturally leads us to the last project, which we might call the *Empirical and Integrative Project*. Unlike the other two, this project is not motivated by questions that arise in various parts of moral theory; it is not primarily normative, nor is it primarily semantic or metaethical. Rather, the methods are descriptive and explanatory, and the aim is to understand the nature of the emotions – in our case, the emotion of disgust.

The project is integrative because understanding the nature of disgust requires the use of a number of different tools drawn from disciplines that make up the cognitive sciences. For instance, it draws on conceptual and theoretic resources drawn from the

following disciplines: the philosophy of mind, including the philosophy of emotions; philosophy of science, specifically ideas in the philosophy of psychology concerning what a proper psychological explanation looks like, and what it is supposed to explain; evolutionary psychology, specifically ideas about cognitive architecture, the structure of minds, and the types of generalities about causal interactions that can take place between different mental states; and gene-culture coevolution, a type of evolutionary thinking that seeks to understand the role of culture in the formation and operation of human psychological capacities.

The project is also empirical in that it draws on information and evidence gathered from a variety of approaches, such as: various branches of experimental psychology, including social, developmental, behavioral economics, and so forth, which use controlled experiments to capture patterns in behavior; cognitive neuropsychology, where researchers are beginning to peak inside the brain and investigate correlations between neural activity, behavior, and other psychological capacities; cultural anthropology, which provides valuable data about cultural variability and universality; and evolutionary biology, which can offer insights drawn from comparing similarities and differences between species.

With respect to disgust, some goals of the Empirical and Integrative Project are to construct a proximate explanation of the psychological mechanisms that underlie the main features of the emotion, including those responsible for production of the emotion and the characteristic features of the response, as well as those responsible for the ability to learn or acquire new elicitors of disgust. Another goal is to produce an ultimate explanation of the evolutionary pressures that gave rise to this emotion in its current

form. Ideally, this component of the project would help illuminate the primary and auxiliary roles that disgust in fact plays in our cognitive economy, including those related to morality.

0.2 Overview

The dissertation is broken up into two parts. The four chapters that make up the first part are devoted to issues in the Empirical and Integrative project. The second part consists of two chapters that use the resources previously developed to illuminate issues in the Metaphysical project, with a particular eye towards idea of projectivism and the sentimentalist tradition.

Part I: The Empirical and Integrative Project

Chapter 1: Towards a Cognitive Theory of Disgust

The first substantive chapter is, in essence, a review of the empirical literature germane to disgust. This is less trivial than it may sound, and this first chapter represents a substantial amount of work. There is no single overarching debate or research program to canvass. The literature *touching on* disgust, however, is dizzyingly large. As a result, the scope of this first chapter is far ranging and highly interdisciplinary. Moreover, beyond an allegedly shared subject matter, this body of research is marked by a striking lack of conceptual unity. Data has been gathered from so many different directions that a compilation of results, presented in an unadulterated fashion, would risk appearing completely piecemeal and disjoint, the conceptual equivalent of a cubist painting that tries to represent its object from every angle at once.

The burden of this chapter, then, is not only to gather together and review all of the relevant research, but also to impose some much needed structure on it. To this end, I

first create what I call the disgust *behavioral profile*. It presents, as near as is possible, just the data points, stripped of whatever overt theoretic framework (Piagettian, Freudian, social constructivist, etc.) that data was originally interpreted through. Moreover, it attempts to specify disgust in purely behavioral terms, so as not to beg any questions about the psychological mechanisms underlying the emotion. The first half of the behavioral profile clarifies the core disgust response and its characteristic features, and points out some of the most prominent downstream effects of that response on other behaviors. The second half attempts to specify all of the different types of elicitors of disgust in as concrete manner as possible, so as to not presuppose anything of theoretical interest.

With the behavioral profile in hand, the chapter goes on to construct a cognitive model of the human disgust system. This is a functional model of the type of cognitive architecture that could account for the behavioral profile, which depicts the different subsystems, features, and mechanisms that make up the human disgust system. It charts out the flow of information between those various subsystems, and associates aspects of disgust behavior with components of the cognitive architecture. The three main divisions it makes are between an execution system, which maintains a database of disgust elictors and produces the constitutive aspects of the emotion itself, an acquisition system, which is responsible for the acquisition of those disgust elicitors that are not innately specified, and the variety of common downstream effects disgust often has on other psychological and behavioral activities.

Chapter 2: Poisons and Parasites: The Evolution of Disgust and the Formation of a Uniquely Human Emotion

The first chapter draws together a substantial array of behavioral data, and ends by positing a *proximate* psychological explanation of that data. The next chapter is devoted to the evolution of disgust, and begins to sketch an *ultimate* explanation of the types of evolutionary forces that gave rise to this emotion, more specifically to the execution system that underlies the response. Essentially, this chapter attempts to answer the question "What is the function of the disgust response?" I argue for what I call the Entanglement thesis. The Entanglement thesis provides an answer to the question that, aside from being compatible with the experimental data, is quite interesting in its own right, specifically in its ability to solve certain puzzles about disgust. According to the Entanglement thesis different components of the disgust execution subsystem themselves have fundamentally different evolutionary etiologies. At the heart of the human disgust system, I claim, are *two* distinguishable mechanisms, each with its own distinct origin and function: one that has to do with diet and the avoidance of toxic foods, and another that has to do with avoiding pathogens, parasites, and the reliable indicators of their presence. Mechanisms evolved to handle each of these problems are present in other animals (the corresponding adaptive problems are not unique to humans) but for a variety of reasons which I draw out in the course of the argument, those mechanisms have merged into a single system only in humans.

Adopting this view suggests immediate answers to many puzzling issues surrounding disgust, most obviously why it has been thought by some researchers to be uniquely human, while others see clear homologues in other primates and animals. Moreover, it offers a plausible account of other much discussed but hitherto unexplained aspects of disgust, including but not limited to: the set of the false positives that trigger

disgust; why conspecifics are so salient to this emotion and why it plays so a prominent role in regulating social interactions; why the defining characteristics of the response form a nomological cluster despite being, prima facie, unrelated; why such a diverse set of entities and objects all trigger this single response; and why certain of those elicitors are universally disgusting. Each of these is discussed in turn, and I show why other accounts of disgust, including the Simple Continuity view and Terror Management Theory, are unable to account for them.

Finally, the chapter ends by offering a preliminary list of the factors that drove these two distinct mechanisms, with their distinct functional and evolutionary trajectories, together into a single deeply integrated system underlying the emotion of disgust in humans. This argument is extended to show why a similar instance of descent with modification did not take place in the cognitive architecture of other animals, even our closest primate cousins.

Chapter 3: Disagreement Over Disgustingness: Variation by Way of Acquisition

This chapter sets aside the execution subsystem, and focuses on the acquisition component of the disgust system, and phenomena associated with that. Evidence increasingly suggests that many disgust elicitors are universal and innately specified, while at the same time it remains just as clear that many elicitors are learned as well. Indeed, preliminary data and much anecdotal evidence suggest that many of the putatively socially acquired elicitors exhibit a pattern of within culture similarity and cross-cultural diversity. Ultimately, we would like to know how this particular process of acquisition occurs, and why – not only why some elicitors are innate while others are

learned, but also why this particular population level pattern of within group similarity and between group differences is found in disgust elicitors.

This chapter develops resources to address these issues. It begins by discussing some of the ways in which a disgust elicitor can be acquired via individual learning. It then gives a proximate explanation of the cognitive mechanisms underlying social acquisition that focuses on emotion expression and recognition. Emotion expression and recognition have been studied at length in their own right, and the fruit of this research is brought to bear in investigating the mechanisms underlying disgust acquisition. These mechanisms themselves have many interesting features, including, most notably, the fact that recognition of disgust (along with most other basic emotions) is universal and often empathic. Recognizing an expression of disgust – a gape face, for instance – often involves *feeling* the emotion of disgust. Production, expression, and recognition of the emotion are all bound together because they all use the same cognitive mechanisms, namely the execution subsystem. I review recent work that supports this, including work in cognitive neuropsychology, that discusses the role of feedback, mimicry and microexpressions in empathic recognition and acquisition.

Finally, I argue for the Empathic Acquisition thesis, which holds that the mechanisms involved in disgust recognition and expression provide a powerful route for the social acquisition of disgust elicitors, in large part due to their fact that recognition is empathic. I conclude by briefly discussing a class of social phenomena that are likely to be influenced by the cognitive mechanisms discussed in this chapter.

Chapter 4: Moral Disgust and Tribal Instincts: A Byproduct Hypothesis

In this chapter I once again take an evolutionary perspective, to develop an ultimate explanation of the mechanisms of social acquisition discussed in chapter 3 and begin refining our understanding of the roles disgust has come to play in regulating social interactions. This sketch appeals to a variety of selective pressures generated by increased sociality and group living.

While the Entanglement thesis defended in Chapter 2 links, disgust to poisons and parasites, disgust is involved in more than food and disease. In order to shed light those roles that outstrip food and disease, and particularly those that are associated with morality, I place disgust in the context of gene culture coevolutionary theory (see Boyd & Richerson 2005). I elaborate on the relevant aspects of this work, which sees humans as being distinctive in the extremely different types of environment they can successfully inhabit, their degree of cooperation or ultrasociality, and their reliance on culture. I particularly emphasize the tribal instincts hypothesis. This corollary of coevolutionary theory maintains that humans came to rely on socially transmitted information to a sufficiently high degree that a core coevolutionary feedback loop was generated, wherein statistical regularities in the cultural and epistemic environment began to exert selective pressures on the innately specified cognitive mechanisms underlying the acquisition and transmission of social information. These selective pressures endowed us with uniquely human "tribal instincts." I separate out three distinct areas in which human psychology manifests such tribal instincts: imitation, social norms, and ethnic boundaries.

Next, I go on to consider how performance of the primary functions of the core disgust system would have been enhanced by the availability of social information, and how the need to better regulate food intake and disease avoidance would have begun

selecting for mechanisms of social transmission and acquisition. In the case of disease avoidance, the interests of any particular individual and the interests of other group members coincide, as contagious infection by any individual would be easily transmitted to any other member of the group. While the same does not hold of food intake, forces associated with kin altruism and inclusive fitness would have selected for mechanisms that allowed parents to signal information to their offspring about what potential foods to avoid.

Having discussed the tribal instincts hypothesis and the social character of disgust in isolation from each other, I go on to begin weaving those pieces together. I argue for the Co-opt thesis, which holds that while retaining most of its core structural features, disgust is involved in all three components of our tribal instincts. The features of core disgust, especially the rigidity of the behavioral response and the open-ended flexibility of the acquisition system, made it a strong candidate to be co-opted when it interacted with the novel conditions produced by the core coevolutionary feedback loop. The features of mimicry and feedback associated with empathic recognition can be explained in part by appeal to the new selective pressures that strongly favored imitation. The role of disgust in many moral judgments can be explained by the core disgust system working in conjunction with a norm psychology that evolved to help coordinate social interactions and produce behaviors that are locally adaptive, given the specific demands of different niches and circumstances. Ethnic boundary markers are often highly emotionally charged, and attitudes and behaviors associated with ethnocentrism, xenophobia and prejudice often follow the logic of disgust, depicting outgroup members not just as wrong or different, but as subhuman, tainted, even contaminating. This is explained by appeal

to an ethnic psychology that evolved to maximize interactions between ingroup members, and that draws on the core disgust system to provide the motivation to avoid members of other tribes.

The picture that emerges is that of a universal but multifaceted cognitive system that is uniquely human in a number of ways. The core system is a kludge, formed when human evolution went down a unique pathway that caused the previously distinct mechanisms underlying taste aversions and disease avoidance to fuse into a single, unified psychological system. That system is unified in the sense that the single, distinctive response pattern is produced whenever the system is activated.

With this account in hand, I return to questions set out at the beginning of the chapter about the relation between disgust and morality. After highlighting some of the difficulties that arise for attempts to demarcate the domain of morality, I show how our account of the link between disgust and morality takes the form by-product hypothesis, and trace out how it is able to explain some of the more puzzling and irrational effects that recent research on disgust and moral judgment has been discovering.

Part II: The Metaphysical Project

The first part focused on the Empirical and Integrative project, and showed that a number of distinct conceptual frameworks are indeed compatible with each other, and that synthesizing them can afford deeper and mutually reinforcing insights into how the mind works. As mentioned above, the unified theory of disgust presented in the first couple of chapters serves as a centerpiece for the later chapters. In the second half of the dissertation, I explore implications of that theory for issues that arise in the Metaphysical Project.

A brief interlude, entitled "Shifting Gears: From the Philosophy of Psychology to Metaphysics," is situated between Chapters 4 and 5 and serves to signal the change in focus. It elaborates on Metaphysical Project, and briefly describes a framework for metaphysical inquiry recently advanced by Alvin Goldman, and within which the final two chapters should be understood.

Chapter 5: Projectivism Psychologized: A Philosophic Idea in Cognitive Scientific Clothing

In this chapter I update the tradition of projectivism, which descends from Hume's observation that the mind has a "propensity to spread itself on external objects" (Treatise 1.3.14). I begin by characterizing that tradition and its enduring allure, and go on to describe some of its historical incarnations, as well as some of the more recent forms it has taken in the wake of the linguistic turn.

Despite that enduring allure, opposition to projectivism has united philosophers with little else in common. I consider what I will call the Master argument against projectivism, which has been advanced, in slightly various guises and with different emphases, by such diverse philosophers as Barry Stroud (1996), Hilary Putnam (1990), John McDowell (1985), and Stephen White (2004), and whose conclusion is that projectivism is (or that many uses of projectivism are) incoherent. I sketch the context in which the argument is usually advanced, trace out its logic, as well as the premises and assumptions on which it relies.

Next, I turn my attention to constructing a notion of projection using the tools of modern psychology, starting with current work linking anthropomorphism to certain properties of our folk psychological capacities, including the implicit operation and

productive output of the autonomous cognitive mechanisms that subserve them. After using anthropomorphism to reconstrue the notion of a projecting mind, I extend the notion to disgust and disgustingness. Here, additional features of the psychology of disgust are of use in illuminating the psychologized account of projectivism, including the fact that it is a kludge, it is multifunctional, and it has a very flexible acquisition system. Such features create an explanatory role that appeal to a projecting mind can easily fill: a nearly unavoidable imperfect fit between response and object in the case of the first two features, and the generation of substantial individual and cultural level variation, in the case of the second.

I conclude that the master argument fails to get any traction on this reconstructed account, and so fails to show it to be incoherent. I end by responding to some of the more obvious objections the account might provoke, and remarking on its wider prospects, scope and limits.

Chapter 6: That's Not Disgusting!: A Critique of Three Views of Disgustingness

This last chapter examines and criticizes three different accounts of the property of disgustingness. It begins by motivating the types of questions and issues that the three accounts serve to answer, and places them in the larger philosophic landscape. Next, it goes on to consider two different functionalist views of disgustingness, each modeled on analogous accounts that have been advanced for color. Before descending into the details of those accounts, though, I give a brief sketch of the history and foundations of the functionalist tradition, in order to better illuminate subtle differences between those two accounts. I advance criticism of each of these account that are based on the empirical work of disgust done in Part I of the dissertation: each is ill equipped to deal with the

cultural and individual variation of disgustingness. Furthermore, the fact that the emotion of disgust is a multifunctional kludge is shown to raise difficulties as well.

The next section moves on to consider a sentimentalist inspired account. Disgust has been especially prevalent in discussions of metaethics recently. It has served as a paradigm example of the type of psychological response in which many metaethical views wish to root our moral nature and capacities, including sensibility and sentimentalist views (see McDowell 1998, Wiggins 1987b, Gibbard 1991, Blackburn 1993, Nichols 2004, D'Arms and Jacobson 2000, 2005). Sentimentalist views descend from Hume and Shaftesbury and see the emotions as playing some crucial role in morality and moral judgments. Many of these theorists agree that a middle ground must be found between robustly realist views such as intuitionism, on the one hand, and such extreme views as moral nihilism or Mackie's error theory, on the other (1977).

After charting out the issues that motivate these types of views, I consider a sophisticated modern variant, D'Arms and Jacobson's (2000, 2005) "fittingness" account of the objects of sentimental responses. I then show how the theory of the emotion developed earlier demonstrates that the disgust response "fits" with few if any of the objects that elicit it. I argue that the proper conclusion to draw about disgustingness is that it is projected onto the world by the mind.

Chapter 1: Towards a Cognitive Theory of Disgust

1.1 Introduction

The emotion of disgust offers an intriguing brew of nature and nurture, the universal and the specific, the innate and the learned. On the one hand, the capacity to be disgusted, together with a small set of things that appear to be universally and innately disgusting, comprises part of the species typical psychological endowment. These are part of human nature, and they do not have to be learned. On the other hand, the variation exhibited in what people can find disgusting shows that nurture has a role to play as well. We learn what to be disgusted by through individual experience, through interacting socially with others, and through the type of education that constitutes the refinement of our moral and aesthetic sensibilities. Due in part to this multidimensional diversity, the emotion of disgust has begun attracting the attention of enough researchers to have become relevant to a variety of debates in different parts of academia, most prominently philosophic debates about metaethics, sentimentalism and response dependence (McDowell 1985, 1987; D'Arms & Jacobson 2000, 2005; Nichols 2004), and empirical moral psychology (Haidt et al. 1993, Haidt et al. 1997, Schnall et al. 2004), but also including a variety of other research projects across the spectrum in psychology.

The recent surge of interest and empirical work on the psychology of disgust has been accompanied by only the mildest convergence in theoretic views, however. Beyond agreement that disgust is a specific type of aversion, a dizzying array of conjectures have been made about its fundamental nature: disgust is a reaction formation, a defense against

or rejection of emotional intimacy (S. Miller 1986, 1993); it is a socially constructed moral emotion of exclusion most closely linked to touch and smell (W. Miller 1997); it is a food-based emotion most closely linked to the mouth (Rozin et al. 2000); it is an innate system evolved to protect us from parasites, germs, and disease (Curtis and Biran 2001); it is, at least in part, a pan-mammalian adaptation that regulates sexual conditioning (Fessler and Navarrete 2003, 2004); it underlies a particular kind of social stigmatization (Kurzban and Leary 2001); it helps in demarcating ethnic boundaries (Boyd and Richerson 2005); it is governed by the laws of sympathetic magic (Nemeroff and Rozin 2000). After only a cursor glance, one might be tempted to wonder whether everyone is talking about the same thing. Closer inspection shows, I believe, that certain of these fragments of theory are compatible with each other, but the fact remains that at this point there is no single received view, accepted by all interested parties. The closest thing to orthodoxy was Paul Rozin's view (Rozin et al. 2000). Even that has come under direct attack from various quarters in the last couple of years, however; see W. Miller (1997), Charash & McKay (2002), Curtis et al. (2004), Fessler & Navarrete (2005) and our next chapter.

A number of factors have lead to the current situation in psychological work on disgust. It is partially due to a trend familiar from other areas of science, namely that in this case data have recently been accumulating faster than theory has been able to keep up. Another part is due to the puzzling and seemingly contradictory nature of disgust itself. But another cause of the proliferation of theory is a feature that is common to most emotions. Like most emotions, disgust is 'level ubiquitous' (De Sousa 1987). Roughly speaking, something interesting can be said about its character from nearly every level of

analysis, from its associated patterns of neural activation to its role in large-scale cultural dynamics, and most points between.[1] From the perspective of a theoretician, this is a particularly exasperating source of confusion, since appreciation of level ubiquity can make it unclear where to even begin in theorizing about the emotions. Moreover, as illustrated by the collection of views just cited, disgust appears to present an especially acute case of this difficulty. For, as reflected by the fragments of theory mentioned above, analyses offered about disgust from different levels of inquiry often seem to have little to do with each other.

This current state of play within psychology is largely what motivates this and the next few chapters of my dissertation. Disgust is puzzling and intriguing in a variety of ways, but despite this – or more likely, *because* of it – no coherent theory has yet emerged to resolve the puzzles or systematically accommodate the data. The aim of these first few chapters is to construct a theory that is able to bring order to the chaos, and which can be brought to bear on the most pressing philosophic debates about locating value in the nature world. As we will see, doing so will require use of conceptual tools drawn from a number of distinct research programs. One of the corollary benefits will be that the resulting theory of disgust can also serve as a case study, showing how diverse conceptual tools can be seamlessly integrated in theory construction.

But before we get there, we need to locate a place to even begin construction of our theory. The standard place to begin such an endeavor would probably be to consider each currently available theory fragment in turn. After subjecting each to criticism and

[1] See Keltner & Haidt (1999) for an exploration of some of the intermediate levels, focusing on the various social functions that emotions perform.

evaluation, we could decide upon the most plausible to defend, supplement, or develop in new directions.

We will not be proceeding in this manner, though. As hinted at above, the fragments of theory that have been put forward so far are tantalizing and frustrating in about equal measure. Their sheer number would make weighing them all against each other burdensome at best, futile at worst. In light of this, I suggest that the best approach is to avoid becoming entangled with the vagaries of those speculations, and begin by returning to the ground floor of what we know, to the facts. Therefore, we will set to the side all theoretic proposals, at least to begin, and instead focus exclusively on the large body of data that has been gathered about disgust. The first step will be to gather those facts together in one place, and construct what I will call the behavior profile of the emotion of disgust. We will conclude by offering a model of the cognitive architecture that begins to explain the facts gathered in the behavioral profile.

1.2 The Behavioral Profile of Disgust

Since it is a compilation of the known facts, the behavioral profile will proceed, in essence, like a review of the empirical literature germane to disgust. Speaking of "the" literature on disgust is a bit misleading, however. Just as there is no single theory, there is also no single overarching debate, experimental paradigm, or research program specifically devoted to this emotion in particular. The amount of empirical work that *touches on* disgust, however, is dizzyingly large, and data is being gathered and reported by researchers from numerous disciplines, with very little overlap by way of shared background assumptions and methodological protocol. As a result, the scope of this first chapter is not only far ranging but also wildly interdisciplinary. Moreover, beyond an

allegedly shared subject matter, this body of research is marked by a striking lack of conceptual unity.

Our burden in compiling this behavioral profile, then, is threefold. First and foremost, we will gather together and review all of the relevant research. Second, we will attempt to present, as near as is possible, just the data points, stripped of whatever overt theoretic framework (Piagettian, Freudian, social constructivist, etc.) those data were interpreted with in the original articles. In so doing, we will try to specify disgust in purely behavioral terms, so as not to beg any questions about theory or the psychological mechanisms underlying the emotion. We cannot remain completely agnostic however, since we need to organize the data in some way. Our third task is to impose some much-needed structure on this otherwise sprawling body of data. In choosing that structure, we will be guided by the structure of disgust itself. The main division we will use to help organize the presentation is between data about the response, on the one hand, and data about the elicitors, on the other.

The first half of the behavioral profile clarifies the core disgust response and its characteristic components, and points out some of the most prominent downstream effects of that response on other behaviors. This section includes neurological data as well; in calling it the behavioral profile, I use "behavior" in a loose sense, and so include data about the way the person's brain "behaves" when she is disgusted. The second half attempts to specify all of the different types of elicitors of disgust in as plain a manner as possible, so as to not presuppose anything of theoretical interest.

Two last caveats: first, we are sketching the capacity as it typically manifests in normal, fully formed, adult human beings. Issues about development and varieties of

malfunction will only be addressed when relevant. Second, the behavioral profile contains only data explicitly about disgust. Other, relevant subject matter (comparative data about primates or conceptual tools borrowed from work on cultural evolution, for instance) will be addressed in later chapters, when relevant.

1.2.1 The Response

Roughly speaking, the response is the way people react once they have detected something that they find disgusting, the pattern of behavior they exhibit when they are disgusted by something. This response has long been thought to be universal, found in all cultures and normally functioning adult humans. Darwin initially provided evidence that all normal, mature humans have the capacity to be disgusted, and that facial expressions of disgust are recognizably the same across cultures (Darwin 1872). Evidence supporting these claims to universality has continued to accumulate even since, and few have found grounds to disagree (see Ekman 1992, Rozin et al. 2000). The exact parameters of "normal" do remain somewhat unclear, however. Among the many deficiencies found in humans raised in extreme isolation is the lack a fully developed disgust response and elicitor set (Malson 1972).

Considerable effort has been dedicated to carefully mapping the different affective, cognitive and behavioral facets of the disgust response. As we shall see, the pattern of behavior making up the response is somewhat idiosyncratic, in that the components of that pattern do not always share any clear thematic unity. In what follows, the properties of the response are broken down into three parts, what we will call the affect program, core disgust, and downstream effects. These parts, and the order of

their presentation, correspond to their relative distance downstream from the initial detection of an elicitor.

1.2.1.1 The Affect Program

The term "affect program" is a conspicuously theoretical notion that wears its commitment to the computational theory of mind on it sleeve. It is taken from psychological research, where it is used to characterize a family of the most basic emotions. In general, affect programs are emotional responses that are complex and highly coordinated. The responses are reflex-like, in that they are often triggered automatically, and have a quick onset and brief duration. Moreover, individual affect programs are triggered by entities and events that have recurring adaptive significance, to which the each particular response is fitted. The historical roots of the conception lie in Darwin, and can be traced through the notion of an innate fixed action pattern used by classical ethologists such as Lorenz and Tinbergen, into its current form in the more contemporary psychological work on emotion done by Ekman, as well as other such as Tompkins and Izard (see Griffiths 2001 for references and discussion). Most of these researchers agree that affect programs are likely to be related to homologous response patterns found in other primates, and to be pancultural amongst humans.

Structurally, an affect program is composed of the a number of parts: a) a trigger or stimuli, which elicits b) a signature behavioral and c) signature physiological response, each of which has its own components, including, most prominently, a characteristic facial expression, and finally d) an attendant qualitative feeling. The response itself is (usually) automatically elicited, and the different elements of that response cluster

together. That is, once an affect program is set off, it automatically triggers not just one or a few of the distinguishable elements of response, but all of them.[2]

The relation of affect programs to emotions in general, especially to higher level or more cognitive emotions, is a tricky one that has been treated at length elsewhere (Griffiths 1997, especially chapters 4 and 9). Paradigm examples of affect programs, however, include anger, fear, joy, sadness, surprise, and, of course, disgust. For the most part, the components of the disgust affect program are easy to identify and separate out. Behaviorally, disgust produces an immediate aversive or withdrawal response, wherein the disgusted subject attempts to distance him or herself from the offending entity. This rejection need not always manifest as moving away, however, but can often result in some other form of getting rid of the offending entity. The associated facial expression of disgust is known as the 'gape face'. It is characterized by a nose wrinkle, extrusion of the tongue and expelling motion of the mouth, and wrinkled upper brow. The gape face mimics the facial movements that precede or accompany actual retching, from which the expression is thought to derive. Like other affect program facial expressions, it is thought to be universal, and universally recognizable as such (Ekman 2003).

In terms of the physiological component, triggering disgust causes a slight drop in temperature, and it is the only affect program marked by a drop in heart rate (albeit a minor one), rather than a rise (Ekman et al. 1993). In addition, disgust increases salivation and gastrointestinal activity. Together with heart rate deceleration these components have been taken to indicate activation of the parasympathetic nervous

[2] As has been noted by both Ekman (2003) and Griffiths (1997), affect programs bear a striking enough resemblance to Fodorian modules as to perhaps constitute being an instance (Fodor 1983). The extent of the overlap is unclear, however, due in no small part to the fact that the notion of a module has become increasingly vexed in recent years (see Fodor 2000).

system, which plays a broadly inhibitory role in the functioning of an organism (Levenson 1992).

Finally, the qualitative component of the disgust affect program is the all too familiar experience of revulsion, and the feeling and physiological concomitants of nausea (Ekman 1992). In fact, this connection to the digestive system, suggested by the feelings of nausea, the increase in salivation, and so forth, has been further elucidated by brain imaging techniques. Evidence gathered using fMRI technology links disgust to the anterior insular cortex, which is thought to be involved in gustatory responses on independent grounds (Phillips et al. 1997). Indeed, it is often called the 'gustatory cortex', and is active in the processing of offensive tastes in both humans and other primates (Kinomura 1994, Rolls 1994). This connection to the gustatory cortex marks disgust as having a neural substrate distinct from other emotions, which are more closely associated with amygdala.

1.2.1.2 Core Disgust

The emotion of disgust outstrips the affect program, however. While disgust appears to bear all of the distinguishing characteristics of affect programs in general, there is more to it; the cluster of elements that comprise the entire disgust response cannot be captured using only the resources of the affect program template. Another set of features that are slightly less reflexive and more cognitive in nature is also produced. Following Rozin's terminology (Rozin et al. 2000) we will call this set of elements of the disgust response *core disgust*. The three central features of core disgust are a sense of oral incorporation, a sense of offensiveness, and contamination sensitivity.

The sense of *oral incorporation* is perhaps most closely related to the affect program. The disgust response generates aversion via many of the same bodily systems employed in digestion and food consumption; nausea, increased salivation, activation of the gustatory cortex and gastrointestinal system are centered on the mouth and digestive system. These components accompany *all* disgust reactions, even those induced by entities that are not potential food or have little to do with eating or the mouth. This fact, however, can create a strong *cognitive* association between the mouth and oral functioning, on the one hand, and all elicitors of disgust, whether or not they have anything to do with the mouth, on the other hand.

Indeed, research has found the aversion produced by disgusting entities can be made more intense by considering those entities as food, or as present in the mouth (Rozin et al. 1995). Feces are disgusting enough; imagining eating them is downright vile. Other studies less directly address this issue, but are obviously relevant to the sense of oral incorporation. For instance, electrical stimulation of the anterior sector of the insula, conducted during neurosurgery, evoked nausea and the sensation of being sick, as well as the feeling that the stomach was moving up and down that often precedes vomiting (Penfield & Faulk 1955). More recently implanted depth electrodes have been used to electronically stimulate the anterior insula, which produced sensations in the throat and mouth that were difficult to stand (Krolak-Salmon et al. 2003; see also Wicker et al. 2003).

A more cognitive, sustained sense of *offensiveness* is also evoked by any entity that induces disgust; those entities are thereafter treated and thought about in a certain characteristic way. Offensiveness is a specific type of aversion, and it goes beyond

merely pulling away or expelling an item from the mouth. The very presence and proximity of disgusting entities is upsetting; they tend to capture attention, and are both memorable and difficult to ignore; they are perceived as unclean, somehow dirty, tainted, or impure; and agents seek to distance themselves from those entities, either by fleeing or by removing the entities from their immediate vicinity. Such behavior is often accompanied by a motivation to cleanse or purify oneself. When the elicitor is more symbolic than concrete, subjects will often try and distance themselves from disgusting ideas or perpetrators symbolically as well, or expel what is offensive by whatever symbolic means seem appropriate (see Rozin et al. 2000).

This feature of core disgust more clearly outstrips the affect program template. Where the affect programs are reflex-like, marked not only by their quick onset but brief duration, this sense of offensiveness is more enduring. Once some particular entity has triggered the disgust system and has thus been marked as offensive, that person tends to treat that entity as such indefinitely, all other things being equal. She continues to be offended by the item well after the reflexive withdrawal is complete, or she stops gaping at it.

Finally, *contamination sensitivity* refers to the fact that once an item is marked as disgusting and offensive, the item can infect other items with its offensiveness; it can contaminate otherwise pure and un-disgusting entities. The means of contamination can vary, but the most common means are via perceived physical contact, or a known history of physical contact or close physical proximity (see Nemeroff & Rozin 2000, Siegal 1988, Siegal & Share 1990).

Contamination sensitivity has a few strikingly idiosyncratic properties. First, contamination is a means by which disgustingness is *transmitted* from one entity to another. Contaminated entities are thereby disgusting, and so induce disgust and are treated in the same way as other disgusting entities. Importantly, there need not be any perceivable residue left by the "source" entity on the contaminated or "receiving" entity in order for an agent to continue treating the receiving entity as if it were contaminated by the source entity. Entities so contaminated are then treated as disgusting, and thus elicit all of the features of the disgust response, including contamination sensitivity – they are treated as being able to transmit their own offensiveness to still other entities.

Second, contamination sensitivity is *elicitor neutral*. Any elicitor of disgust, regardless of the actual nature of the elicitor, or which disgust "domain" it is from (physical, social, moral, or otherwise), has contamination potency of the same basic sort. If any item is disgusting, it is thereby considered contaminating, and can transmit its disgustingness to other entities in the same way.[3]

Much of the experimental work of Paul Rozin and his colleagues investigates these properties, while also documenting the surprising strength and ubiquity of contamination sensitivity. Some experiments demonstrate subjects' refusal to drink juice that has come into contact with disgusting items, such a cockroach or human hair. Others use the same format to show that there need be no actual physical contamination to trigger contamination sensitivity. In some cases, subjects refuse to drink juice that has come in contact with demonstrably clean, *uncontaminated* entities, such as cockroach which has been chemically sterilize, or a brand new comb or flyswatter, removed from

[3] This appears to be a special case of the more general consistency of the disgust response across different domains. See Borg et al. (forthcoming) for confirming brain imaging data.

the plastic in front of the subjects (see Rozin et al. 1985, Rozin et al. 1986, Rozin et al. 1989). Still other experiments measure subjects' increasing aversion to clean sweaters that have been contaminated by their histories. For instance one sweater used in these experiments was new, while another was a sweater that was laundered after it was worn once by a perfectly healthy stranger. Subjects showed greater reluctance to put on the used sweater, which they considered somehow contaminated despite the fact that it put through the laundry. Even more interestingly, subjects' contamination sensitivity increased substantially when they were told that the previous owner of the sweater had experienced a misfortune such as a leg amputation, had a disease such as tuberculosis, or was a convicted murderer. Most aversive of all was a sweater that once belonged to Adolph Hitler (Rozin et al. 1994). This is a vivid demonstration of elicitor neutrality, the striking fact that disgusting entities are *all* contaminating, regardless of the character of whatever elicited the particular episode of disgust. In other words, Hitler's moral disgustingness is at least as contaminating as the more concrete disgustingness of a cockroach, or a human hair.

Third, there is an important *asymmetry* between disgustingness and non-disgustingness when it comes to contamination potency. This asymmetry is often talked about in terms of purity. Consider the fact that it is far easier for something pure to be contaminated than it is to purify something that is already contaminated. Or to illustrate: a single drop of sewage can spoil an entire jug of wine, but a single drop of wine doesn't much help in purifying a jug of sewage. Evidence suggests that common sensical observations such as these are on the right track, and that this asymmetry is indeed a cross-cultural feature of the disgust response. For instance, neither American nor Hindu

Indian children (4-8yrs) regarded potential purifiers (addition of color to the juice, boiling, or mother taking a sip, indicating it to be okay) as effective at rendering the contaminated substance "clean" or pure again (Hejmadi et al. 2004, see also Nemeroff & Rozin 2000 of the cross cultural ubiquity of the "laws of sympathetic magic").

In sum, contamination sensitivity, even more obviously than the sense of oral-incorporation and offensiveness, does not comfortably fit anywhere in the affect program template. It is a more cognitive feature of the response, rather than a brute physiological or reflexive one. The sensitivity to a disgusting entity's contamination potency endures long beyond the immediate reaction it produces, but production of that sensitivity is part of the response nevertheless. Thus, in addition to the reflexive features grouped together in the affect program, the three properties of core disgust are also part of the homeostatic cluster that makes up the disgust response.

It is also worth emphasizing that though "offensive" and "contaminating" are properties often ascribed to *items* that trigger disgust, a sense of offensiveness and contamination sensitivity and the patterns of behavior and inference associated with them, in the sense discussed here, are parts of the *response* to such items. Indeed, one of the most insidious aspects of disgust is that once an item triggers it, that item is thereby treated *as if* it were offensive and contaminating – whether or not it is genuinely offensive (if there is such a thing) or objectively contaminating (which there certainly is). In this sense, then, it is part of the disgust response that the properties of offensiveness and contamination potency are *projected onto* whatever elicits it.

1.2.1.3 Downstream Effects

We should begin this section by explaining what is meant by downstream effects. In order to do this, we need to step back and reflect on what we are doing. In compiling this behavioral profile, we are using experimental data to sketch the contours of a behavioral capacity, namely the capacity to be disgusted and all that that entails. Schematically speaking, the behavioral profile is the explanandum, it is the set of data that a theory of disgust will explain. The resultant theory will constitute a psychological explanation. As such, it will appeal to the structure and functioning of psychological entities, namely underlying cognitive mechanisms, in order to explain the behavioral capacity, the typical patterns of behavior and inference in question.

The first step in giving a psychological explanation, then, is to clearly characterize the capacity being explained. However, as many commentators on psychological methodology have pointed out, individuating a capacity is far from trivial (see for instance Cummins 2000, or Prinz 2004, chapter 1). The immediately relevant upshot of this difficulty is that there is not always a straightforward way to distinguish between one capacity and another, between the essential features of some capacity its downstream effects, the ways that capacity's operation affects other activities, cognitive, behavioral, or otherwise.

This general worry about individuating a capacity and isolating the primary target of explanation can be raised for the emotions, including the particular case of disgust. To deal with this worry we will proceed thus: we will make the assumption that the features of the affect program and core disgust can be treated as the essential features comprising the capacity to be disgusted. The behavioral data to be described in this section, on the other hand, can be separated off and relegated to the status downstream effects of disgust

proper. This means that rather than essential components of the capacity to be disgusted, these data reveal the systematic effects of disgust on other, distinct cognitive and behavioral capacities.

Several considerations justify the assumption that the elements of the affect program and core disgust comprise the capacity to be disgusted. First, this portion of the response exhibits *consistency*; whenever disgust is induced, whatever the nature of the elicitor and the context, the coordinated response that is produced reliably includes all of the elements of the affect program and core disgust.[4] The straightforward fact that these elements all regularly covary with each other suggests a single capacity gives rise to them, and thus they are all essential features of that capacity. Second, while researchers are unable to agree on much of theoretic substance about disgust, all seem to identify the emotion they are interested in by reference to the features I have gathered together under the headings of the affect program and core disgust. Third, many of the remaining behavioral features that I classify as downstream effects clearly involve the operation of other capacities. Indeed, many of the experiments they are drawn from are explicitly designed to test the effects that inducing disgust will have on other capacities and systems.

Even if these considerations do not firmly establish the division drawn here, we will adopt it as a working hypothesis anyway. The theory developed later, if on the right track, will help to vindicate the assumption on which it was predicated. Assuming this working hypothesis, then, we can move on to investigate the data on downstream effect, which often gives clues to the structure and functioning of the capacity itself. For

[4] This is not to say that elements of the response, or the entire response itself, cannot be voluntary suppressed in certain social contexts, exaggerated in others, or similarly shaped by certain culturally specific norms of expression.

instance, part of the offensiveness of disgusting entities is that once detected, they tend to capture attention, stick in the memory, and increase sensitivity to other potentially disgusting entities. A series of correlational studies reveals *attention* and *memory biases* for disgust elicitors; all else being equal, people pay more attention and better at remembering disgusting things than neutral ones. (Charash & McKay (2002). The results also provided an instance of what is sometimes called mood congruency, the idea that being in a particular mood or emotional state makes one more sensitive to elicitors of that emotion. When primed with disgusting stories beforehand, people paid significantly higher attention and were better at recalling disgusting things than others.

Memory and attention biases are probably related to the fact that disgust also tends to induce a bias towards *information sharing*, making people more likely to tell others about things that disgust them, to pass along cultural items that are associated with disgust. Once again, experiment has provided support to casual observation on this score. In one study (Heath et al. 2001), the focus here was on urban legends: embellished stories about recent, often lurid events, that sometimes contain a grain of truth (but often do not), that are popularly believed to be true, and that spread quickly through a population either way. The study found that subjects were more likely to pass along an urban legend that was disgusting than on that was not, and were more likely to pass along particular urban legends the more disgusting they were. In addition, they found that the more disgusting a story was, i.e. the more disgusting motifs it contained, the more likely it was to show up on a set of urban legend websites.[5] Another study indirectly supporting the existence of an information sharing bias looks at the most prominent etiquette manuals over the last

[5] Disgustingness in the first two experiments was measured by self-report of the subjects, but in the third, web-based experiment disgustingness of urban legends was measured using the Disgust Scale (Haidt et al. 1994).

few centuries (Nichols 2002b). It finds that etiquette norms prohibiting behaviors that are likely to trigger disgust (spitting while at the dinner table) were more likely to be passed down through generations than those that are not (using the wrong fork to eat a salad).

The emotion of course has a powerful phenomenological component, and this gives rise to a proprietary and all too familiar vocabulary (W. Miller 1997). That vocabulary, colorful though it may be, needs no exemplification here.

Some of the most notorious downstream effects of disgust involve the influence it can exert on evaluative judgment about a variety of subject matters, including morality and economic decision making. Least surprising is the fact that disgust can have a negative influence on evaluations, making them harsher and more severe. What is particularly striking about this downstream effect that it is extremely *persistent*, in that it survives through a number of conditions. In the simplest case, the elicitor of disgust and object of evaluation (be it an entity, action, etc.) are one and the same, and the person is in possession of good reasons to support her judgment. In such cases, people make more negative evaluations, and are able to articulate justifications for why they make those judgments.

More eyebrow-raising are cases where the disgust elicitor and object of evaluation are the same, but all reasons offered in support of the negative judgment can be defeated. In such cases, the disgust response again produces a negative evaluation. Moreover, the bald disgust response has a powerful enough effect on judgment that people will continue to endorse their initial negative evaluation even upon reflection. That is, people will maintain their negative judgment of the object of evaluation even when they admit that, by their own lights, they are unable to articulate any good supporting reasons. Jon Haidt

and others, who continue to explore the influence of disgust (and other emotions) on moral judgment (Haidt et al. 1993, Murphy et al. 2000; see also Haidt 2001), have dubbed this phenomenon *moral dumbfounding*: people make persistent moral judgments, but are dumbfounded as to what might justify them. For instance, many subjects held fast in the condemnation of disgust inducing activities such consensual sibling incest or masturbating with a dead chicken, even when they have been convinced that none of the reasons they initially give in support of the judgment are credible (Murphy et al. 2000).

Most unsettling is the fact that disgust, once induced, can negatively affect judgments even when the object of evaluation is *distinct* from the elicitor of disgust. In one rather devious set up, hypnotism and disgust were used to produce negative and relatively more severe judgments of blameless, innocuously described vignette characters. Those who experienced hypnotically induced disgust (triggered by otherwise neutral words in the vignettes) were unable to pinpoint why they disliked the characters in question, but judged them to be suspicious and untrustworthy nonetheless (Wheatley & Haidt 2005). Subjects were hypnotized to feel a flash of disgust at arbitrarily chosen words, such as 'often' or 'take'. They were then given a series of vignettes describing moral transgressions, each of which they were to rate for morally wrongness and disgustingness. Across the board, ratings were more severe when disgust was induced. Subjects in whom disgust had been hypnotically triggered gave more severe ratings, both for moral wrongness and disgustingness, and for both moral transgressions that involved disgusting actions (cousin incest and eating one's dog) and those that did not (a politician who takes bribes, an ambulance chasing lawyer). Most interesting was subjects' reactions to the following neutral vignette, which describes no moral transgression, nor

hints at anything wrong or disgusting: "Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He [tries to take/often picks] topics that appeal to both professors and students in order to stimulate discussion." Use of the disgust inducing word in the vignette, however, increased judgments of disgustingness and moral wrongness by factors of roughly 10 and 6, respectively. Subjects maintained their unfavorable judgment of Dan, despite their complete lack of justification for it, dubbing him a "popularity-seeking snob" who "just seems like he's up to something" (page 783).

Moreover, this type of persistent downstream effect on evaluative judgment appears to be produced even in less devious experimental set ups, where people *realize* the source of their disgust and object of judgment are distinct, and when they *know* the two have little or nothing to do with each other. So-called "carryover effects" have been found to affect judgments and decisions on a wide variety of subject matters, including moral judgments. In one particularly vivid example, subjects were first given a survey to determine how sensitive they are to bodily signals when deliberating, and how much affect influences their decision making process. Those who scored high on this survey again made more severe moral judgments when they had been subjected to an "extraneous" disgust prime that putatively had nothing to do with the vignettes they were asked to rate (Schnall et al. forthcoming). In one experiment, disgust was induced by having the subjects fill out the Disgust Scale (Haidt et al. 1994). In the second disgust was primed by having the subjects rate the vignettes at a desk that was intentionally made filthy: "An old chair with a torn and dirty cushion was placed in front of a desk that had various stains, and was sticky. On the desk there was a transparent plastic cup with the

dried up contents of a smoothie, and a pen that was chewed up. Next to the desk was a trash can overflowing with garbage such as greasy pizza boxes and dirty-looking tissues." Again, for the subjects sensitive to their own body signals, even moral judgments were more severe when disgust was induced. This was true for vignettes that described disgusting moral violations, and more surprisingly, it was also true for judgments of the moral violations that had nothing to do with disgust.

Disgust has been shown to affect other sorts of cognition in similar ways. One study found that disgust has an impact on risk aversion, at least in women (Fessler et al. 2004). This experiment was inspired by evolutionary considerations, and rather than focus on disgust, it looked at the downstream effects of multiple emotions on various types of reasoning. They found that "extraneously" induced disgust reduced risk-taking behavior in women subjects. In another, subjects who were primed with disgust in a "normatively unrelated" setting (watching a four minute scene involving a filthy toilet from the film *Trainspotting*) failed to exhibit what behavioral economists know as the endowment effect (Lerner et al. 2004). The endowment effect is the much-studied phenomenon wherein the minimum price subjects are willing to sell an object for, after it has been given to them (is endowed to them), is significantly greater than the maximum price they would be willing to buy it for in the first place. Lerner et al. showed that when disgust had been induced in subjects beforehand, the asymmetry was eliminated; the prices subjects consented to in the selling and buying conditions were roughly identical. Moreover, both prices were lower than either the buying or selling conditions in the neutral condition, when no emotion was primed, or in the sadness condition, in which the endowment effect was reversed. Note that in both these cases, participants are fully

aware that the object of their evaluation and the elicitor of their disgust are distinct. Nevertheless, disgust demonstrably and systematically alters their reasoning in both cases, again exemplifying the extreme persistence of disgust's negative downstream influence on evaluative judgments.[6]

1.2.2 The Elicitors

The other half of the disgust behavioral profile is the set of those things upstream from disgust responses, namely the elicitors. While the makeup of the disgust response exhibits consistency across all of the things that induce it, the pool of elicitors is remarkably diverse. Many have speculated about the nature of disgustingness, and the thread that all disgusting things have in common. For instance, theorists have hypothesized that triggers of disgust are pollutants, or matter out of place (Douglas 1966) or they are reminders of death and our animal nature (Rozin et al. 2000). We will reframe from adjudicating between attempts to capture what all disgust elicitors have in common. We avoid this for methodological purposes, but also because these attempts rely on a dubious assumption, namely that disgust elicitors all share some property above and beyond triggering disgust. Rather than argue against that assumption here, however, we will again stay as close as possible to the facts, and confine our efforts to listing the known elicitors as specifically and concretely as possible.

One potential pitfall we should flag is the liability to confuse the projective character of the response with actual properties of the elicitors. For the sake of clarity, it is worth pointing out that however natural or correct it sounds, saying something like

---

[6] Though the results aren't as straightforward or easily interpretable, other studies have suggested another link between disgust and economic decision-making. For instance, using an fMRI on subjects participating in an ultimatum game, Sanfey et al. (2003) found heightened activity in the anterior insula (the gustatory cortex associated with disgust) in reaction to unfair offers, and found that increased activity in the same area predicted whether a subject was likely to reject an offer.

"disgusting things induce disgust" is not of much help, since it is question begging on the face of it. Given that the response includes elements like contamination sensitivity, sense of offensiveness, and feelings of nausea, it is no more help to merely say that contaminating things, offensive things, or nauseating things induce disgust. Rather, these better describe the effects that elicitors have on people who are disgusted by them (though, for instance, some things that are *treated* as contaminating are *actually* contaminating as well). Part of the disgust response is that one experiences nausea, and that contamination potency and offensiveness are projected onto the elicitors via the patterns of behavior with which they are treated.

Finally, we should single out the gape face, which is an elicitor of a slightly different sort than those discussed below. Recognition of the gape face (and other aspects of disgust expression) can be said to elicit disgust because recognition is often *empathic*: it involves the recognizer actually experiencing the emotion they recognize being expressed by another. Moreover, voluntarily making a gape face (or performing any single element of the disgust response) often triggers the entire cluster of elements, and produces an experience of the emotion in the person making the gape face. Rather than include them below, however, we will save the discussion of these aspects of disgust for Chapter 3.

1.2.2.1 Some Candidate Universals

There is an unarguable affinity between disgust and various sorts of organic materials. Hence, at the most concrete end of the spectrum of elicitors are what Rozin and others have suggested as the best candidates to be universals: feces, vomit, urine, and sexual fluids (Rozin et al. 2000, Angyal 1941). Equally likely candidates are corpses and

signs of organic decay, which are also some of the most potent elicitors of disgust (Haidt et al. 1994). Bodily orifices – and via contamination, things that come in contact with bodily orifices – are likewise powerful and potentially universal elicitors (Rozin et al. 1995). More generally, artificial orifices, or breaches of physical bodies such as cuts, gashes, lesions or open sores (in Rozin's terms, violations of the ideal body envelope) are further candidates for disgust universals. These can trigger disgust either if they occur to one's own body – in which case they might also cause pain – or in someone else's. In this sense disgust appears universally sensitive to the boundaries of organic bodies, and in many cases is activated when those boundaries have been, or are in danger of being, breached.

Bodily boundaries are operative in triggering disgust not only when they are in danger of being violated, however. Items and substances once within those boundaries, which were once inside or part of the body, but that then exit or are detached from the body, constitute a related class of potentially universal elicitors of disgust. Severed limbs and externalized innards, either your own or those once belonging to others, fit this description; so, too, do the waste products mentioned above. Other classic examples of this are blood and saliva. Swallowing the saliva that is currently in your mouth is innocuous; even imagining drinking a glass of spit, even if it is (or was?) your own, is revolting. The blood in your or anyone else's veins is fine; an unchecked nosebleed or spurting artery is disgusting. Fingernails and hair are other good examples of body parts that are innocent enough when still attached, but become aversive once they become detached – especially when they are in danger of reentering via the mouth (W. Miller 1997).

In this sense, disgust not only polices the bodily boundaries, but is also the enforcer of a "No Reentry" policy; anything that exits or becomes detached elicits it.[7] These elicitors also involve physical bodies, their structure, composition, and the ways they can breakdown, and as such, look to be plausible candidates for universals. Aside from the intuitive plausibility and persuasive preliminary evidence, these also all involve organic features of bodies that are themselves human universals, and by and large do not vary with of age, physical environment, culture, or ethnicity.

Finally, reliable marks of disease and parasitic infection provide another plausible set of disgust universals. Signs of disease include those exhibited by other humans who are infected, as well as environmental signs that reliably indicate the presence of infectious agents. Indeed, knowledge that some person is infected with disease can make that person disgusting to others, even those others are fully aware that that disease in question is not contagious (Rozin et al. 1992). While many have noted the associations between disgust and infection, recent experimental work has marshaled overwhelming evidence supporting the connection between the two, gathering input from over 40,000 subjects from 165 countries. In one study that used web-based techniques, subjects rated a range of photographic stimuli on how disgusting they were, and found a similar pattern from subjects the world over: photos of objects indicating potential disease were judged more disgusting than similar images that lacked disease typical signs (Curtis et al. 2004). In another, people from a variety of cultures were asked what disgusts them, and researchers then ran a statistical comparison between the reported elicitors and a list of

[7] See also Fessler and Haley (forthcoming) for more on disgust and the bodily perimeter.

infectious diseases. They found that "for every disease, one or more elicitors of disgust was [*sic*] mentioned as playing an important role" (Curtis & Biran 2001).

1.2.2.2 Some Common Themes

One of the better-known features of disgust is that it exhibits substantial individual and cross-cultural variability. Thus, the remaining types of elicitors exhibit more variation than those listed above, and so make less plausible candidates for universals. Within the evident variability, however, some common themes stand out. For instance, disgust is often induced not just by people who exhibit reliable indicators of disease, but by a more general set of morphological irregularities and phenotypic abnormalities. "Phenotypic abnormality" appears to be a theme with considerable room for variation, and has been hypothesized to include, in some cases, people who are disfigured, handicapped, obese, elderly, and even members of an outgroup who are unfamiliar or foreign looking.

One meta-analysis looks at data from a variety of previous studies, and argues that this triggering of the disgust system underlies the aversion some feel towards the disfigured and handicapped. They further speculate that the same holds true of aversion to the elderly, obese, and perhaps any conspecifics that deviate too far from the cultural ideal in morphology (Park et al. 2003). A study by this same group suggests that heightened disgust sensitivity correlates with xenophobia; unfamiliar or foreign looking people can be disgust elicitors as well (Faulkner et al. 2004). At this point, it is not completely clear what this amounts to, or what 'unfamiliar' or 'foreign looking' denote, but plausible candidates include characteristics, morphological, physiognomy or otherwise, that mark people as members of an outgroup or different ethnicity.

Brain imaging techniques have offered support for the link between disgust, ethnocentrism and prejudice towards outgroup members. While filling in the details, they have revealed a particularly troubling aspect of the phenomenon as well: a correlation between disgust and dehumanization. Subjects were shown pictures of members of a variety of social groups. In those cases of prejudice where disgust was the accompanying emotion, and only in those cases, the medial prefrontal cortex (mPFC) failed to activate (Harris & Fiske 2006).[8] The mPFC is the brain area associated, on independent grounds, with higher-level social interactions with other people, and is thought to underlie theory of mind and the attribution of agency. This suggests not only that disgust is elicited by members of certain outgroups, but that it is elicited particularly by those outgroup members who are dehumanized, not even thought of as people or agents.

Food is another common theme in disgust elicitors, as Rozin has emphasized for many years. Though all cultures deem some foods disgusting (and, on the other side of the coin, embrace foods that other cultures find distasteful or disgusting), the particular foods falling into these categories vary from location to location, and from culture to culture. Moreover, these foods are often considered disgusting for conceptual or symbolic reasons. Rozin and his colleagues also point out that disgust is distinct from mere inappropriateness, i.e. not eating something because it is considered inedible, or mere distaste, i.e. rejecting something merely because it tastes bad (Fallon and Rozin 1983, Rozin et al. 2000). Indeed, disgusting foods are distinct from distasteful or inappropriate ones in that they are treated as offensive and contaminating, and hence are unlikely to get into the mouth to be tasted in the first place. Additionally, foods that

[8] See also Cottrell & Neuberg (2005) for evidence that different outgroups produce prejudicial attitudes associated with different emotions.

caused gastro-intestinal sickness when they were previously ingested by an individual, or were merely correlated with such illnesses when previously ingested, become elicitors of disgust for that individual (see Bernstein 1999 for an overview on taste aversion).

There appear to be biases for which foods might be disgusting, however. One particularly prominent theme found in the distribution of disgust inducing foods over different cultures is meat of various sorts. In light of the association to physical bodies, it is not altogether surprising that meat is a common elicitor of disgust, and in a comparison of food taboos across 78 different cultures, Fessler and Navarrete (2003) found that meat consumption is more often regulated or restricted than consumption of other foods, in large part due to the role of disgust.

Some living animals, and not just their products or corpses, are liable to elicit disgust as well. These include many "creepy-crawlies", and animals that are highly associated with disease, decay, and death, which are perhaps linked to disgust mainly in virtue of this association. Flies, maggots, worms, rats, and cockroaches are obvious examples. Others, which are in fact parasitic on humans, include lice, fleas, and ticks. In addition, Davey and colleagues have identified another group of animals that humans often find aversive, and whose aversion is driven by disgust. It includes slugs, snails, caterpillars, as well as animals that can be dangerous to humans, but are not predators: snakes, and especially spiders (see, for instance, Davey et al. 1992, Webb and Davey 1993, Ware et al. 1994).

Another common theme in disgust elicitors is sex and reproduction. For instance, menstrual blood is more disgusting than other types (Rozin et al. 2000). Disgust is also triggered not just by sex-associated fluids, but also by many of the sexual activities that

produce them. The most discussed instance of this is incest (Fessler & Navarrete 2004, Lieberman et al. 2002, Westermarck 1921), but other types of deviant sexual activities evoke disgust as well. While "deviant sex" induces disgust in most everywhere, what counts as deviant is, to some extent, dictated by particular cultures, and can differ from one culture to the next. For instance, homosexuality might be considered deviant and disgusting, as in many parts of the U.S., or might be perfectly acceptable, as in other parts of the U.S., or ancient Greece (see Haidt & Hersh 2001). As in the case of food, there appears to be constraints on the variance that is possible here, as more extreme varieties of deviance such as bestiality and necrophilia are more likely to be deemed disgusting.[9]

A final theme in disgust elicitors includes activities, and their perpetrators, that involve breaking some social norm. While the particular activities that fall into this set vary from culture to culture, all cultures appear to find some social transgressions disgusting. Some such transgressions are probably disgust inducing because the social norm being violated regulates an activity that involves an antecedently disgusting substance or activity. In other words, violation of social norms governing, for instance, the locally correct way to deal with corpses or dispose of fecal matter, how to properly prepare food, or conduct oneself at the dinner table or in the bedroom, are likely to induce disgust merely in virtue of the subject matter being regulated. A variety of data indirectly support this. One study focusing on etiquette norms used excellent examples of elicitors of this sort, which include norms against picking one's nose in public, or spitting into a glass of water and then taking a sip while at a dinner party (Nichols 2002a). Another found that many different languages have words that roughly translate to "disgust", and

[9] Fessler & Navarrete (2003) document a further wrinkle in the link between disgust and sex. They show that sensitivity to sexual elicitors of disgust, but only sexual elicitors, heightens during certain phases of women's menstrual cycles, peaking when they are most able to conceive.

that are likewise applied to social activities of these sorts (Haidt et al. 1997). The meat taboos mentioned above (Fessler and Navarrete (2003) constitute more examples of this type, as do many of the vignette's used to explore the effect of disgust on evaluative judgments, such as consensual brother sister incest or masturbating with a chicken carcass (Haidt et al. 1993).

Again, while norm violations of this sort constitute a common theme of elicitors of disgust, the particular prescriptions and proscriptions of such taboos and purity norms can vary from culture to culture. The variation can be found along a number of dimensions, including their specificity as well as in the importance and centrality of such norms to the local socio-moral code (see Shweder et al. 1997, Rozin et al. 1999).

However, violations of norms having little or nothing to do with the types of elicitors mentioned above can also trigger disgust. The common theme here is quite abstract, but appears to be that someone flouting a particularly central social norm or violating a defining value can induce the disgust of other members of the cultural ingroup. For instance, the Hopi value the environment, the Greeks prized self-control, the Japanese place a high value on duty and social cohesion, and the American self-image assigns importance to egalitarianism, personal integrity, and rugged individualism. Social activities that violate these have been found to elicit disgust in each culture, respectively. Likewise in the U.S., Republicans and Democrats define themselves against and in opposition to each other; those in the opposite party, who espouse the opposing ideology, are liable to elicit disgust.

One particularly interesting cross cultural study that looked at, among other things, disgust and the violation of defining social norms (Haidt et al. 1997). Examples

of these in the United States, listed when subjects where asked the open-ended question of what they find disgusting, included acts of racism, hypocrisy, violations of important social relationships, dishonest politicians and opposing political attitudes. In their own words, "Lawyers who chase ambulances are disgusting. People who abandon their elderly parents are disgusting. Liberals say that conservatives are disgusting. Conservatives say that welfare cheaters are disgusting" (page 116). Japanese participants mentioned, along with other, more universal disgust elicitors, situations where they failed to meet their own standards, when they felt shamed or abused by others, and when they felt others had failed to meet their needs or expectations. Ancient Greeks felt disgust towards those who flouted social norms and conventions due to lack of self-control, or those whose transgressions were unaccompanied by shame; they were barbarous, inhumane (Parker 1983). Perhaps the most telling description of this class of disgust inducing activity comes from the Hopi, whose specific elicitors include disregard for the environment and any form of aggression: "Anything that would be deviant to Hopi teachings and belief could be seen as disgusting to some degree" (quoted from Haidt et al. 1997, page 120).

1.2.3 Shaping the Theory: A Pair of Constraints

We began by remarking on the reasons that disgust has become a focal point of research in philosophy and psychology, and noted that for all the interesting data being gathered, sophisticated theory construction has lagged behind. Rather than begin by examining the various conjectures that have been made, we got back to the facts, and constructed a clean set of data that any theory of disgust needs to explain.

We conclude this section by pointing out that, in addition to the brute facts, the character of the behavioral profile and the proliferation of theoretic conjectures can also offer guidance in theory construction. For, it is not unreasonable to want an adequate theory of disgust to explain not just the data, but provide some insight as to why so many different but plausible things can be said about this emotion. Seeing the embarrassment of riches this way points to a pair of key desiderata:

> *The Unity of the Response*: The characteristic disgust *response* is comprised of a number of distinct features. These features form a homeostatic cluster: they occur together as a package, and regardless of what triggers disgust on any particular occasion, once it is triggered the production of one element of the cluster is regularly accompanied by the production of the others. What accounts for the clustering of this idiosyncratic set of features? Why have these particular cognitive, behavioral and physiological elements merged into a single, unified, and apparently universally human, response type?
>
> *The Diversity of Elicitors*: A wide and surprisingly diverse range of *elicitors* trigger disgust, ranging along one dimension from the very concrete to the very abstract, along another from the universal to the culturally and individually specific, and along another from the brutely physical and inert to the highly social and interpersonal. What accounts for the pairing of such a large variety of triggering conditions to this one specific type of response?

Next, we begin constructing such a theory by offering a model of the type of cognitive architecture than might give rise to disgust and can satisfy these constraints.

1.3 A Psychological Model

We now begin the task of constructing a theory of disgust. Since this first step takes the form of a psychological model of the cognitive architecture and proximate mechanisms that underlie disgust, we should begin with some brief remarks about the general character of these sorts of theoretic tools, and how they are understood to do their explanatory work.

1.3.1 General Background Assumptions

First, the type of explanation offered here is a proximate explanation, rather than an ultimate one. The distinction between proximate and ultimate explanation was first brought to prominence in the context of biology, but it can be brought to bear for psychological explanations as well (Mayr 1961, Ariew 2003, Barkow et al. 1992). In the psychological case, a proximate explanation explains behavior by reference to stimuli in the immediate environment and the structure and functioning of internal psychological mechanisms. Ultimate explanations, alternatively, are evolutionary, and thus historical. In the psychological case, various behaviors, and often the character of the underlying psychological mechanisms themselves, are explained by appeal to the selective pressures that helped form them, and the adaptive problems they evolved in response to. Although it is important not to confuse one for the other, ultimate and proximate explanations often complement each other. Both types are required for a complete theory of disgust. Accordingly, we here give a proximate explanation; the next chapter will consider an ultimate one.

Next, we should comment on the explanatory relations between the behavioral profile and the psychological model. With the behavioral profile, we have sketched the contours of a particular behavioral capacity, namely the capacity to be disgusted. This capacity is comprised of patterns of behavior, broadly construed, and is thus described in behavioral terms. To explain that capacity, our (proximate) theory of disgust will be couched in psychological terms, and the description of the model will make reference to the likes of cognitive architecture and cognitive mechanisms. It is the operation of these psychological entities posited in the model that produces, and thus explains, the patterns of behavior described in the behavioral profile, and that comprise the capacity to be

disgusted. Succinctly put: the capacity is the explanandum, and the psychological model the explanans (Cummins 2000).

Although they are the standard conventions, it is also best to be explicit about what the various elements of our model are being used to represent. The model depicts a cognitive architecture. The term "cognitive architecture" simply provides a graphic way of talking about the structure of minds, understood as generalities about the types of causal interactions that can take place between different mental states. It is a functional level model, and attempts to account for the various types of data compiled in the behavioral profile with a cognitive architecture composed of different but interlocking subsystems and mechanisms. It does this by charting out the flow of information between those various subsystems, and associates various aspects of disgust behavior with corresponding components of the cognitive architecture the help produce them.

It is depicted as a boxology. Different "boxes" represent functionally distinct components of the mind, and the arrows represent causal relations between them.[10] Each box stores a propriety body of information that leads to the production of the patterns of behavior with which it is associated.

Most boxologies, including our model, are founded on the twin doctrines of functionalism and the computational theory of mind. Roughly speaking, functionalism is the ontological thesis that mental states and properties are functional properties, whose identity conditions are determined by their functional role and specified mainly in relation to other mental states and the behaviors they cause, or could cause. The computational theory of mind is based on the computer analogy, the idea that the

[10] This style of explanation owes much to expositors of homuncular functionalism (Fodor 1968, Dennett 1978, Cummins 1983).

relationship between the brain and the mind very much like the relationship between the hardware of a computer and the programs it runs. The computational theory of mind supplements functionalism's ontological picture with the more specific claim that mental processes are computational processes performed on mental representations.[11]

1.3.2 The Disgust System

We now turn to the model. It is a first pass pitched at a fairly high level of abstraction, but it is in this abstraction from detail that much of the model's utility resides, as one purpose it serves is to impose a map on an otherwise sprawling body of evidence.

The model divides the cognitive architecture into three main parts. The first part is an acquisition subsystem. This component of the disgust system is responsible for acquiring those disgust elicitors that are not innately specified. The significance of the acquisition subsystem stems from the need to account for the cultural and individual variation found in disgust elicitors. That variation indicates many elicitors are acquired from the environment, either form individual experience or social learning. Acquisition of either sort is to be explained by appeal to the performance of underlying cognitive mechanisms in the acquisition subsystem.

The second is an execution subsystem. In addition to producing the core elements of the disgust response, as mapped out in the behavioral profile, the execution subsystem

---

[11] The idea is that the program can be expressed in a formal language of mental representations, which are individuated by their syntactic structure. The computational, and thus mental, processes that operate on those mental representations are sensitive only to that syntactic structure (as opposed to their content, for instance). On this general picture, mental representations will eventually be paired with neurological states by a function that maps members of one set to members of the other. The discovery and construction of this function is the general goal of empirical cognitive psychology. The preference for a syntactically structured language as the formalism of choice is based on the hope that the syntactic relations between the mental representations will mimic the causal relations between the neurological states they encode (Fodor 1975, Stich 1983). Also see Schiffer (1981) for a much more detailed discussion of the relation between talk about "boxes" and talk about functional roles and mental representations.

also maintains a database of elicitors. That database contains representations of items and entities that trigger the disgust response when they are detected in the environment.

The third part of the cognitive architecture depicted in the model does not represent a component the disgust system proper, but shows the variety of other psychological and behavioral activities upon which disgust has been found to have systematic downstream effects.

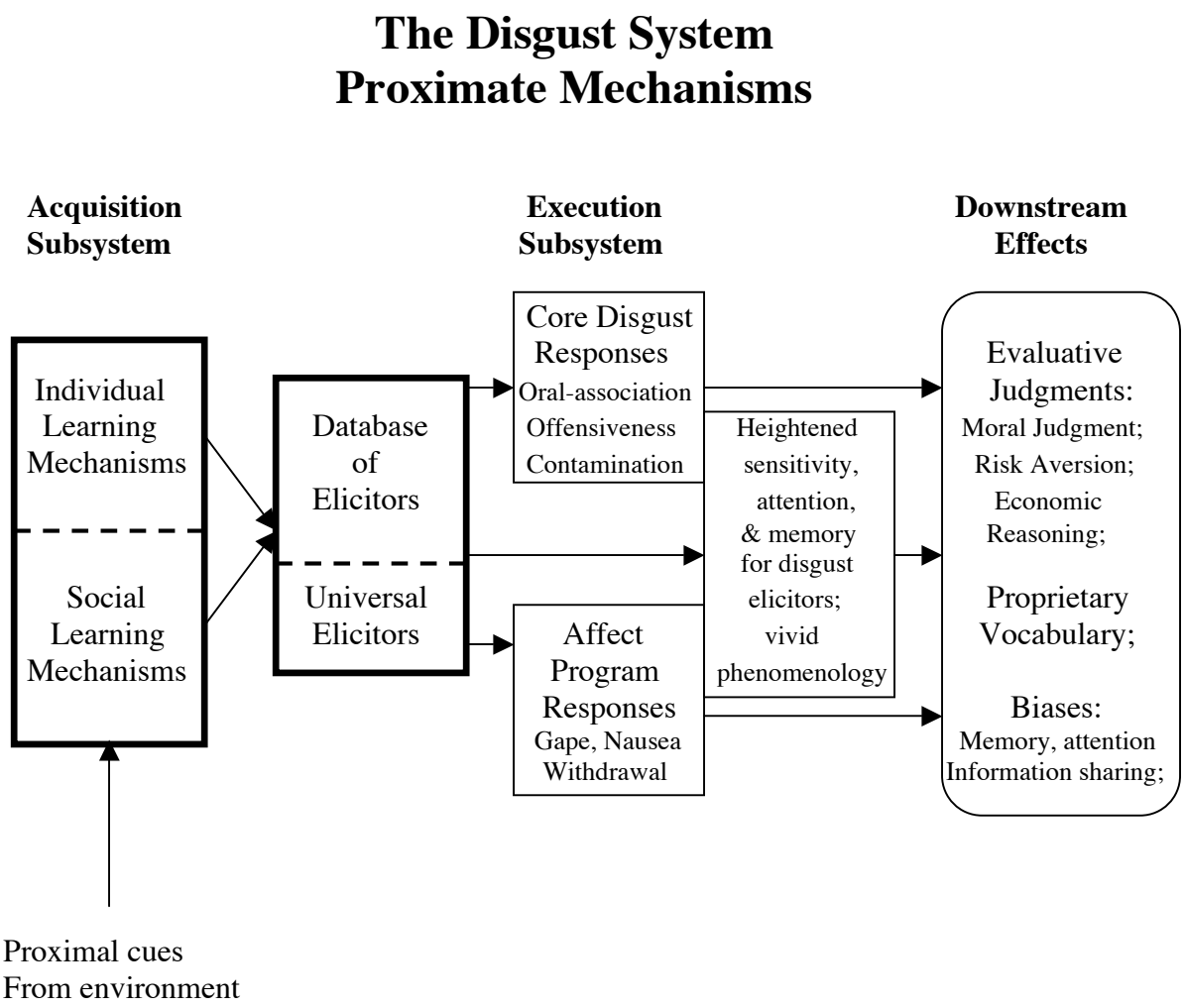


Figure 1.1
A functional level model of interlocking mechanisms that comprise the human disgust system. The arrows represent causal links between the various mechanisms.

It will be useful to carefully walk through the model, beginning on the left with the acquisition subsystem. This is represented as containing a number of distinct, independently operating cognitive mechanisms. A division is made between mechanisms that rely on individual learning and those that rely on social learning. The broad function of mechanisms on both sides of this divide is to pick up on relevant cues and patterns from the surrounding environment, social or otherwise, and infer from them new contents for the disgust database. Mechanisms are divided by the kinds of proximal cues to which they are sensitive: the former of these involve acquisition of elicitors via direct

interaction with or experience of them, unmediated by social transmission or other people. The later involve acquisition of elicitors from other people, via imitation, explicit learning or other forms of social transmission. Individual and social learning, in this respect, probably represent poles along a continuum rather than a sharp functional distinction. The distinction is clear enough to be useful for organizational purposes, however; thus the dotted rather than solid line.

As it stands, the acquisition subsystem allows for a plurality of acquisition routes and mechanisms, but remains agnostic as to their number and individual character. It is likewise agnostic with respect to how restricted or open are the conditions under which an elicitor may be acquired via each mechanism; some mechanisms may deliver a new elicitor only under very specific circumstances, while others may be able to perform in a wider variety of cases. The model is committed, however, to the fact that at least some of the mechanisms of acquisition are innately specified, and that those are likely to exhibit many of the characteristics associated with innate cognitive mechanisms, such as domain specificity, automaticity, and stable developmental trajectory.

This pluralism accounts for one sense in which the disgust acquisition subsystem itself is quite flexible. For, what unites all of the diverse acquisition mechanisms is that they are all able to deliver new elicitors to the disgust execution subsystem. There they are encoded in the database, which is able to receive new elicitors from a number of different acquisition mechanisms.

This brings us to the disgust execution subsystem. Moving from the left, the first component is the database of elicitors. It is depicted as a box, which "contains" a functionally distinguished set of representations of those entities or activities which

trigger disgust when detected in a person's environment, vividly described or imagined, and so on. More specifically, when an item represented in the database is detected, this leads to activation of the full suite of affective, behavioral, and cognitive components that make up the disgust response.

The database is divided into two sections by a dotted line that separates elicitors by their source, rather than their function. Again, once represented in the database, all elicitors lead to the same disgust response, regardless of how they got there, be it individual or social acquisition, or innate specification. On one side are those elicitors acquired from experience. On the other are the innately specified, universal elicitors. Additionally, this later side may contain other types of innately specified information about the sorts of things that trigger disgust. More specifically, the model reserves this spot for innately specified information that might take forms *other* than representations of specific elicitors. Such information might be in the form of constraints, biases or more general guidelines. Information represented in formats such as these may also interact with information drawn from the environment or may require information from the environment to activate, complete, shape, or edit it in some way.

Interlocking with this database are more integrated cognitive mechanisms that produce the characteristic features of the disgust response. There are two distinct mechanisms in the execution subsystem. One corresponds to the affect program, and gives rise to associated elements of the response like the gape face, nausea and quick withdrawal. The other corresponds to core disgust, and gives rise to the associated elements of the response such as oral-association, offensiveness, and contamination.

These mechanisms and the database make up the execution subsystem, whose broad function is to produce the characteristic behavioral features of the disgust response whenever one of the elicitors in the database is detected, vividly described or imagined, and so on. These architectural components of the execution subsystem are assumed to be innate and universal amongst normal, mature humans, and this assumption is again justified by appeal to the fact that the disgust response is pan-cultural. Moreover, the operation of the execution subsystem is largely automatic and involuntary when triggered, acting as a psychological reflex.

By producing the disgust response, the execution subsystem also generates the sorts of downstream effects on other behavioral and psychological activities. These effects appear systematic in some individual cases, but less coherent and tightly integrated than the central features of the response. Thus, the model simply illustrates that these appear to be causally preceded by the activation of the execution subsystem. Other than suggesting a broad pluralism, the model remains agnostic as to the types of mechanisms involved in those downstream effects, or the nature of their interaction with the mechanisms of the disgust system; some of the most striking effects from the behavioral profile are depicted in the model.

Also depicted in the model is mechanism or collection of coordinated mechanisms that increase sensitivity, attention, and memory specifically to disgust elicitors, which were also mentioned in the behavioral profile. Along with a vivid phenomenology, these appear to be more immediate and consistent downstream effects than the others, and are shown closer to the execution subsystem accordingly.

1.4. Conclusion

In providing a proximate explanation of the core features of human disgust system, the model presented here constitutes the first component of our theory of disgust. Moreover, it makes some headway in satisfying our pair of constraints. Recall what they are:

> *The Unity of the Response*: The characteristic disgust *response* is comprised of a number of distinct features. These features form a homeostatic cluster: they occur together as a package, and regardless of what triggers disgust on any particular occasion, once it is triggered the production of one element of the cluster is regularly accompanied by the production of the others. What accounts for the clustering of this idiosyncratic set of features? Why have these particular cognitive, behavioral and physiological elements merged into a single, unified, and apparently universally human, response type?
>
> *The Diversity of Elicitors*: A wide and surprisingly diverse range of *elicitors* trigger disgust, ranging along one dimension from the very concrete to the very abstract, along another from the universal to the culturally and individually specific, and along another from the brutely physical and inert to the highly social and interpersonal. What accounts for the pairing of such a large variety of triggering conditions to this one specific type of response?

Our psychological model addresses the Unity of the Response desideratum by showing that the features of the behavioral response cluster because the proximate cognitive mechanisms underlying the elements of that response are interlocking, and tightly integrated. It addresses the Diversity of the Elicitors desideratum by positing a number of acquisition mechanisms, which, despite differences in the conditions and types of proximal cues they are sensitive to, all function to deliver new elicitors to the disgust database. Thus, different people, exposed to different cultural conditions and with unique individual histories, end up being disgusted by different entities and activities.

Of course, while our model makes some progress, neither desideratum has been completely met. Indeed, the solutions offered by the model can seem to be more restatements of the respective problems than satisfying solutions. This is to be expected,

however, since our theory of disgust not yet complete, either. The next step to providing a more satisfying solution requires us to supplement our proximate explanation with an ultimate one, that looks to the evolutionary origins of the disgust system. We turn to this in the next chapter.

Chapter 2: Poisons and Parasites: The Evolution of Disgust & the Formation of a Uniquely Human Emotion

2.1 Introduction: A Puzzle about Disgust

Let us begin this chapter by posing a few comparative questions. First: is the emotion of disgust found only in human beings? This question is interesting not only for the insight an answer might shed on human nature, but also because different theorists working on the emotions have come to divergent views regarding it. On the one hand, a group of prominent researchers who have focused on disgust in particular answer this question in the affirmative. On the view they recommend, disgust is "a very old (though uniquely human) rejection system" (Haidt, Rozin, McCauley, and Imada 1997), that "is absent in nonhuman primates, yet extremely frequent and probably universal in contemporary humans" (Rozin, Haidt, and McCauley 2000). Proponents of this view are impressed by a number of distinctive features of disgust that they have uncovered in their work, including its decidedly cognitive, symbolic and conceptual character, the role it plays in regulating human social interactions, its wide cultural variation, and its link with a plurality of domains, including morality. Additionally, they note that despite the confidence of some, others profess an inability to actually identify anything that fits the description of disgust in other animals (Chevalier-Skolnikoff 1973, see also Morris et al. 2007[12]).

Additionally, they give an argument that suggests why disgust might be unique to humans. The motivation for the argument comes from the work of the cultural

[12] Morris et al. argue that contrary to received wisdom, there is strong evidence that several so-called "secondary emotions," particularly jealousy, can be found in non-primate mammals such as dogs and horses. Despite this liberal stance towards emotion possession in other species, however, they find little evidence for disgust in those same non-primate mammals.

anthropologist Ernest Becker (1973), author of *The Denial of Death*. Becker, like Nietzsche and Freud before him, assigned great import to the fact that humans, alone amongst the animals, must psychologically confront the knowledge of their own impending deaths. He argued that recognition of our own mortality and eventual death induces existential anxieties, and in extreme cases, even terror. Feelings and attitudes such as these represent an adaptive threat; they can be at worse paralyzing, but even in milder cases can stifle or disrupt normal, fitness enhancing behavior. Building on this idea, Rozin and his colleagues maintain that via a process of cultural evolution, conceptual and symbolic disgust now mainly serves to guard against such paralyzing and fitness reducing thoughts, repressing anything that reminds us that we are animals, and are thus mortal. This, in turn, helps explain why only humans have disgust: "Only human animals know they are to die, and only humans need to repress this threat (Rozin et al. 2000). Following the literature, I'll call this view *Terror Management Theory*.

Another set of factors pulls in a different direction, however. Consider a second comparative question: are their homologies[13] of disgust in primates and other animals? Some researchers on the emotions have thought so, and on their view it would be surprising if there were not homologies of disgust in all sorts of other animals (Ekman 2003, Griffiths 1997, Darwin 1872). While those sympathetic to this view tend to focus on the family of basic emotions or emotions in general, rather than disgust in particular,

[13] Homologies (the term is taken from evolutionary theory) are similar traits or systems whose similarities can be traced to a shared ancestry. For instance, dolphin fins and human hands are homologies – similarities in the bone structures between the two can be traced to an evolutionarily recent common ancestor (despite the fact that they now serve different functions). Homologies are often contrasted with analogies, similar traits or systems whose similarities reflect convergent evolution. Similarities in the structure of the eyes of a human and the eyes of a giant squid, or in the structure of the wings of a bat and the wings of a butterfly, are not derived from recent common ancestry, but from shared function (see Dennett 1995, pages 136-138 for a brief, non-technical discussion).

their confidence in this assertion is bolstered by a number of specific considerations. These include the presence of clear homologies of other basic emotions such as anger and fear in primates and other mammals. Moreover those holding this view often see disgust as serving to monitor food intake and protect against ingested toxins. They thus point to the presence of something approximating the gape face (the characteristic facial expression associated with disgust) in primates, and the existence of acquired taste aversions in many other animals. Perhaps more than anything else, though, they emphasize the broad evolutionary continuity that exists between humans and primates to support their contention. Accordingly, I will call this view the *Simple Continuity View*.

At first blush these two views appear to be opposed to each other. If they are incompatible, we would like to know which is the correct one. One may not be forced to take sides, however, as there are other stances to take with respect to the issue. There may be an irenic conclusion that could be endorsed, holding that each view is partially correct when understood properly. On the other hand, it could also be the case that neither is correct, and both should be rejected.

In what follows, I will argue for this third option. In order to say why, however, I must first motivate and defend the alternative view that I favor. Once that has been done, we will briefly return to the puzzle we started with, and show why both the Simple Continuity View and the Terror Management Terror should be rejected as ultimate explanations and accounts of the fundamental nature of disgust.

2.2 The Entanglement Thesis

Here, in short, is the hypothesis to be defended and elaborated upon below: underlying disgust are two distinct cognitive mechanisms that became functionally

integrated with each other in the face of selective pressures faced by early humans. Thus was a single emotion formed via natural selection, whose character was shaped by features of both mechanisms and the adaptive problems they were designed to solve. While homologies with similar features and functions to each individual mechanism can be found in primates and other animals, only in humans have these two mechanisms become functionally integrated, and thus only in humans do we find this particular emotion. Since we are speculating that two mechanisms became so entangled as to form a single emotion, we will call this the *Entanglement thesis*.

In order to make the case for this hypothesis, we will build on our earlier work. Whereas the previous chapter offered a proximate explanation of the facts compiled in the behavioral profile, in this section we proceed on the assumption that that model is by and large correct, and offer a set of ultimate explanations for some of the mechanisms posited therein. In addition to describing the adaptive problems that gave rise to the distinct mechanisms underlying our capacity for disgust, we will also consider a hypothesis about the conditions and evolutionary pressures that drove those mechanisms together, reshaping and fusing them into a single, integrated system.

According to the proximate model offered in the previous chapter, the two distinct mechanisms (aside from the database) that underlie the disgust response are associated with the affect program and core disgust, respectively. Below we will see that each of these mechanisms is not only behaviorally but also evolutionarily distinct, and the elements associated with each mechanism can be traced back to their evolutionary past. That is, each is evolutionarily ancient: the mechanisms and problems they evolved to mitigate originate far back in human phylogeny, and mechanisms homologous to each

can be found in other species. However, each has followed a quite different evolutionary trajectory: each initially arose to perform a different function. This remains the case despite the fact that they have become deeply intertwined in humans, and apparently, only in humans.

One mechanism, associated with the affect program, evolved as an adaptive response to the ingestion of toxins and harmful substances. The other, associated with core disgust, evolved as an adaptive response to the presence of disease and parasites in the broader physical and social environment.

2.2.1 Food Intake: The Omnivore's Dilemma, Acquired Taste Aversions and the Garcia Effect

The mechanism underlying the affect program is closely linked to digestion, and evolved specifically to regulate food intake and protect the gut against ingested substances that are poisonous, toxic or otherwise harmful. It was designed to expel substances entering or likely to enter the gastro-intestinal system via the mouth, and has been call a "food rejection system" (Darwin 1872, Rozin et al. 2000).

Rozin's work on disgust emphasizes the relevant adaptive problem, which he calls "the omnivore's dilemma". All species that are "nutrition generalists" face this dilemma, given that some potential foods are more nutritious than others, while still others are detrimental. The problem itself is quite simple: the organism must eat, but it must be selective in what it consumes, because many things that are edible are harmful when ingested. Rozin distinguishes a number of ways in which this problem might be mitigated: simple distaste prevents some foods from being consumed based on their sensory properties, namely because they taste bad (bitter, sour, etc.), while disgust can

prevent some substances, including potential foods, from even being tasted in the first place.

One especially well-known way to navigate problems raised by the omnivore's dilemma is provided by acquired taste aversions. These provide a way to narrow down culinary options by implementing a "once bitten, twice shy" rule – a type of food that has induced sickness in the past is avoided in the future. In humans, this variety of "shyness" manifests as a characteristic aversion to the offending food type that bears many of the same characteristics of the disgust affect program.

Much interest in acquired taste aversions focuses on the proprietary mechanism of individual learning. The features of this form of learning were first systematically investigated in rats (Garcia 1974), but similar effects have since been found in an astounding number of other animals, ranging from garden slugs through primates to humans (Bernstein 1999). While some tastes, such as sourness or bitterness, are innately aversive, an aversion to many specific foods must be learned. The learning mechanisms associated with Garcia effects require only a single trial to acquire an aversion to a new type of food. If consumption of a particular food is accompanied by gastro-intestinal stress, even as far as 12 hours after consumption, an aversion to that food is developed.

While it appears that the stress must be specifically gastro-intestinal in order for a taste aversion to be acquired, the stress can be caused by a number of sources. The most obvious source, of course, is ingested food that is itself poisonous or toxic, or which carries other foreign substances with it, such as parasites or other pathogens (more on this below). Moreover, these aversions are more likely to be acquired for foods with strong tastes and pungent smells. The salience of such foods is instrumental in the development

of false positives, which are not uncommon to this mechanism. Aversions may form for sharp tasting or pungent smelling foods even when the gastro-intestinal stress accompanying or following their ingestion is not directly caused by that food, but some something completely unrelated. This is called the "Garcia effect"; the taste aversions themselves are sometimes called "Garcia aversions".

Comparative evidence suggests that the system underlying acquired taste aversions is evolutionarily quite ancient: studies show that similar systems are found in a great many other animals, including those phylogenetically distant from humans. The system would have been highly adaptive in the past, as it is now, and is quite specific in what it applies to. Aversions form mainly for foods that have been ingested at least once, and are more likely to be elicited by the taste and smell of food items than by their other properties. The function of the system, as Garcia and many others have speculated, was specifically to protect the gastro-intestinal system from direct harm.

Perhaps most relevant, though, is the means by which aversion is generated once a taste aversion is acquired. The specific response utilizes many of the same systems involved in ingestion, and often results in feelings of nausea. This, of course, serves immediately to deter an organism from consuming the substance in question. In this connection to nausea and ingestion, the link between taste aversions and human disgust is most manifest. For in humans, the behavioral elements of the disgust response, specifically those associated with the affect program (as well as the sense of oral incorporation noted in core disgust), almost all involve bodily systems also associated with food and the digestive system. For instance, the characteristic facial expression, the gape, involves movements associated with oral expulsion, as well as the constricting of

the nasal passages used to smell food. This facial expression clearly mimics the patterns of muscular contraction in the mouth, nose and face that are involved during the actual behavior of retching. Extreme disgust can result in outright vomiting, but even in milder cases it includes not just quick withdrawal but the physiological element of nausea. In distancing oneself from the offending items, quick withdrawal serves to lessen the intensity of those smells that can trigger nausea.

All of these elements of the disgust affect program bear the mark of a system designed to help deal with the omnivore's dilemma and help monitor food intake. Moreover, the fact that food itself is one of the common themes in the elicitor set only strengthens the conclusion human disgust and taste aversions are deeply linked. Hence, the ultimate explanation of the properties of the affect program is thus to be found in the evolutionary logic of taste aversions. We have one component of the Entanglement thesis in place: one of the two main mechanisms comprising the execution side of the human disgust system is a modern version of the evolutionarily ancient taste aversion system.

2.2.2 Disease and Parasite Avoidance

This leaves the mechanism associated with core disgust, whose response features include oral association, offensiveness and contamination sensitivity. According to our hypothesis, this mechanism was shaped by the adaptive problems raised by disease causing pathogens, and the evolutionary arms race underlying the struggle between parasites and their hosts. It evolved to provide one way to protect against infection from pathogens and parasites, namely by avoiding them. Unlike the affect program, this mechanism is not specific to ingestion, but serves to prevent against coming into any sort

of close physical proximity with infectious agents. This involves avoiding not only visible pathogens and parasites, but also places, substances and other organisms that might be harboring them. The capacity has been described as intuitive microbiology (Pinker 1997).

Though there is little discussion of parasites or infectious disease in psychology (as opposed to psychiatric or neurological disease), the prospect of infection presents a nearly ubiquitous set of adaptive problems. A parasite is any organism that grows on, feeds on, or exploits the resources of another – its host – but that contributes nothing to the host's survival. Given that parasites drain resources without making any contribution in return, the adaptive problems they raise for potential hosts are not only ubiquitous but fairly simple: hosts need to avoid and protect against them, and eliminate parasites once they become infected by them.

Accordingly, natural selection has endowed potential hosts with a series of defense mechanisms against pathogens and other parasites. Within the body, immune systems equipped with an arsenal of antibodies wage wonderfully complicated cellular level warfare on viruses and bacteria. Skin is an external protective membrane that, among other things, provides a defensive barrier against parasites infiltrating the body in the first place. Many animals engage in hygienic behaviors that minimize the likelihood of infection, such as grooming, cleansing, or bathing.

That is not all, however. Natural selection has also endowed potential hosts with capacities designed to help them avoid pathogens and parasites in the first place. These capacities can monitor a wide range of potential sources of infection, and operate by making organisms sensitive to signs of parasites in the environment, and especially to the

proximal cues of parasitic infection in their conspecifics. Such cues can be general, like as the smell of organic rot and decay, or more specific, such as particular aberrant types of appearance or behavior of others. This includes especially salient irregularities in appearance such as lesions, sores, or discoloration, or disruptions of bilateral symmetry. In general, parasite avoidance capacities are predominantly sensitive to any phenotypic abnormalities, or deviations from the healthy phenotypic norms.

Kurzban and Leary (2001) also do a good job of articulating this adaptive problem, and in so doing point out another important feature of pathogen and parasite infection:

> "Because parasites specialize in exploiting the particular biochemical makeup of their hosts, transmission of parasites is most likely between biologically similar organisms. So from the point of view of parasite avoidance, a good strategy is to avoid those who are most similar to oneself, namely conspecifics and members of closely related species."
>
> (Kurzban and Leary 2001, page 196)

In addition to the emphasis on conspecifics, they also point out another form of phenotypic abnormality that is relevant, namely types of behavior that is might indicate infection. In extreme cases, parasites can hijack an organism's behavioral control system, causing it to engage in otherwise abnormal behaviors that specifically help spread the parasite to other hosts (the increased aggressiveness found in canines with rabies provides a good example of this).

Such capacities are likely to have other properties as well. There is good reason to think that some of those cues that trigger avoidance would be innately specified in humans, not only because the system is evolutionarily ancient, but because some signs of the presence of parasites or infection are likely to be both species-specific and universal. Ancestral humans who avoided those cues in the past would be more likely to live long

enough to produce offspring, and organisms manifesting those cues would be less likely to attract a mate.

There is also good reason to think that organisms sensitive to the dynamics of parasitic infection, i.e. the fact that infected substances and conspecifics can be contagious, and can thus pass on their infection, would be more fit as well. They would avoid not just the substances and conspecifics that manifested the telltale signs, but other substances, items, or conspecifics that came in contact or proximity with those infected. Finally, there is good reason to think that the system would be more prone to false positives than false negatives, since there is significantly greater cost – infection and possible death – in mistaking an actual source of infection as clean than in mistakenly avoiding uncontaminated items, places, or conspecifics. "Better safe than sorry," is the appropriate guiding logic.

Given this general description of the adaptive problem and the character of the capacity we would expect to have evolved in response to it, it should be no surprise how widespread capacities of this sort seem to be in the animal kingdom. Evidence of parasite avoidance has been found for a variety of species, ranging from tadpoles and three-spined sticklebacks, to Eastern bluebirds and red-winged blackbirds, to primates such as lemurs, baboons, and chimpanzees (see Kurzban and Leary for some discussion and references). The evidence of such a capacity in humans is even more impressive, though it is not always appreciated as such, and has never brought together in one place.

Disgust has nearly all of the properties we would expect of a parasite avoidance mechanism, given the contours of the adaptive problem that it evolved in response to. More specifically, the behavioral features of core disgust, especially offensiveness and

sensitivity to contamination potency, fit the description almost perfectly. These constitute just the type of behavior we would expect in response to carriers of infectious, potentially transmittable parasites and diseases. Thus, the other component of the execution subsystem is a parasite and pathogen avoidance mechanism.

In fact, the conclusion that human parasite and pathogen avoidance is subserved by disgust is nearly inescapable when one recalls more details from the behavioral profile, including the wide range of prima facie unrelated behaviors and entities that fall in its actual domain. For instance, there is strong evidence that reliable signs of disease are universal elicitors of disgust. Disgust, rather than fear, underlies aversion to non-predatory animals whose threat to humans takes a less direct form than brute bodily harm. This includes some animals that are poisonous, such as snakes and spiders, but mostly animals commonly associated with decay and disease transmission, such as rats, flies, worms and maggots.

Some of Rozin's most striking findings show how the human disgust system is prone to be activated by false positives, including such memorable instances as pooh-shaped chocolate, rubber vomit, and juice stirred with a sterilized cockroach (see Rozin et al. 1986, Rozin et al. 1989). Indeed, this propensity for false positives is a general feature of the system. As noted in the behavioral profile, another common theme in disgust elicitors is that specific types of 'phenotypic abnormality' that do not result from parasitic infection, such as being elderly, disfigured, or handicapped, can trigger disgust nevertheless. AIDS suffers elicit aversion to physical contact and fear of contamination even in those who know the disease is not communicable by mere proximity or touch (Rozin et al. 1992).

Other previously puzzling features of disgust also fall into place once we realize its function in parasite avoidance. Together, eating and sex constitute two of the most basic evolutionary imperatives. Both behaviors are ineliminable ingredients of evolutionary success, but both involve crossing bodily perimeters at various points, and thus leave the participants highly vulnerable to infection. Hence, disgust's apparent role in monitoring the boundaries of the entire body (rather than only the mouth) makes much more sense in light of its connection to infectious disease. Moreover, eating and procreating are specific activities that open up those boundaries. They are highly salient to disgust both because they are unavoidable and because they are two of the most potent vectors of disease transmission.

All of these elements of core disgust bear the mark of disease and parasite avoidance. With this, we now have the second component of the Entanglement thesis in place: the ultimate explanation of the properties of the core disgust component of the human disgust system, as well as an explanation for many of the innate and universal elicitors of disgust, is to be found in the evolutionary logic of parasite and pathogen avoidance.

2.3 Descent with Modification

Mother Nature is a tinkerer, and the human disgust system bears many marks of her tinkering. Before getting into the specifics of that tinkering, however, it is worth reemphasizing that underlying disgust are two integrated but originally *different* mechanisms. Comparison with other species suggests that while each is individually present in, they are often still functionally distinct. Moreover, while both mechanisms are evolutionarily ancient, they have very different evolutionary trajectories. One of

them is specific to ingestion and the gastro-intestinal system, and serves to prevent the oral intake of any sort of substance that has once been harmful to the gut, be that harm due to poison, pathogen, or whatever else might cause upheaval to the stomach. The other mechanism is sensitive to a much wider range of factors. It serves to prevent close physical proximity to any potential sources of infection, rather than only those that might target the gastro-intestinal system.

Supporting this is the fact that the two mechanisms appear to follow different developmental schedules in the course of human ontogeny. The gape face, sensitivity to the facial expressions of caregivers, and other aspects of the affect program are present very near to birth, while at least one major aspect of the parasite avoidance mechanism, namely contamination sensitivity, does not emerge until significantly later. Research suggests that children respond to the facial expressions of caregivers by the time they are a year old (Bandura 1992). While all appear to agree that contamination sensitivity has a later onset, there is controversy on its exact schedule: some studies mark it 4-8 years (Rozin et al. 1986, Rozin et al. 1985, Fallon et al. 1984), while others mark it at 2 ½ to 3 years (Siegal & Share 1990).

According to the Entanglement thesis, these two cognitive mechanisms must have become functionally integrated with one another at some point in human evolutionary history. As such, the human disgust system appears to have been shaped in important ways by the evolutionary process of *descent with modification*. Roughly speaking, a trait (character, system, etc.) undergoes descent with modification when selection pressures gradually alter its structure from one generation to the next. A trait subject to this process will slowly morph over evolutionary time, so that when they appear in different

generations, instances of the trait will exhibit slight differences, but also an underlying similarity. Traits resulting from descent with modification can be found all over the evolutionary spectrum, but most recognized cases involve the modification of physical traits or phenotypic characters.

Disgust, by contrast, presents a case in which natural selection modified the psychological structure of human minds. Through a number of generations, two mechanisms were gradually modified, combined, and integrated to the point where activation of one automatically triggered the activation of the other. This resulted in the formation of the cluster of elements comprising the disgust response: offensiveness, contamination sensitivity, nausea, withdrawal, and a gape face. The reaction to potential carriers of parasitic infection came to involve nausea, and gaping of the mouth. Alternatively, acquired taste aversions turned the offending food not just inedible but offensive, and contaminating.[14] Thus, through the modification of human cognitive architecture, the execution component depicted in our model was formed, and the unified disgust response came to be.

2.3.1 Factors Leading to Entanglement

A natural question to ask at this point is: what factors might have been instrumental in causing these two systems to coalesce into their modern human form as they descended through earlier hominid generations? No doubt such factors were many and subtle, but a few stand out as likely playing a pivotal role. First, prior to the influence of any novel selective pressures, there was a non-trivial degree of *antecedent functional overlap* between the two mechanisms. Mitigating the respective adaptive

[14] Though this is consistent with common sense and anecdotal evidence, I know of no experimental data telling one way or the other on it. As such this constitutes a novel prediction made by our theory of disgust, and which might be used to test the entanglement thesis.

problems associated with each mechanism required the production of aversion of some sort. As noted earlier, food is a major vector for disease transmission, and thus already likely to be salient to a parasite and pathogen avoidance mechanism. Spoiled or decaying food not only smells bad and causes gastro-intestinal upheaval, but is also more likely to carry pathogens and parasites as well – indeed, the pathogens that spoil the food are often the same pathogens that cause the subsequent upheaval. Thus, given the respective adaptive problems each was designed to solve, the two mechanisms were probably good candidates for functional integration to begin with.

Second, there is good reason to think that the significance of this functional overlap between the two mechanisms was amplified by major changes in the diets of ancestral humans. Most important of these changes was an increase in their level of *meat consumption*, either via hunting or scavenging (Leakey 1994). This, in turn, brought with it an increased vulnerability to infection, from both more and novel parasites. The expansion of diet introduced more exposure to disease and parasites due to more frequent proximity to both dead animals and other scavengers. Jon Haidt makes the point nicely in the following passage:

> "During the evolutionary transition in which our ancestors' brains expanded greatly, so did their production of tools and weapons, and so did their consumption of meat (Leakey 1994). … But when early humans went for meat, including scavenging the carcasses left by other predators, they exposed themselves to a galaxy of new microbes and parasites, most of which are contagious -- they spread by contact." (Haidt 2006, Chapter 9)

It is noteworthy that the expansion of diet to include more meat would not have introduced any completely novel adaptive problems to ancestral humans. As noted above, contagious diseases, infections, and parasites are ubiquitous in nature, and capacities to protect against them are found in a variety of other animals. To be

evolutionarily successful, early humans were likely no different. In light of this, the disease and parasite avoidance system was probably not originally *generated* by the expansion of diet to include more meat. Rather, it is more likely that a shift in emphasis was brought about in the nexus of selective pressures relating to disease and parasites. This shift subtly effected a case of descent with modification, driving the relevant avoidance mechanism (which was almost certainly formed by this point in evolutionary history) to play an even more pronounced role in screening potential foods and in shaping practices surrounding food consumption.

A final factor that likely contributed to the fusion of these two mechanisms has to do with the advantages gained by being able to *transmit information* between conspecifics. In the case of humans, emotional facial expressions are not just mere symptoms or functionless by-products of some internal state, but serve to signal information to others. In the case of disgust, what has become a signal of potential for infection and contamination is a prima facie unrelated expression, the *gape*: the facial movements that accompany the expulsion of food from the mouth.

The gape face has another feature that suits it to the purpose of signaling, however, namely, it is easily recognizable. Given the properties of core disgust and the elements of the disgust response we have identified as stemming from parasite avoidance, ancestral versions of that system probably did not have a distinctive, easily decodable component of its behavioral repertoire that could act as a recognizable signal. The gape face, however, could easily have been recruited to serve as one. As the two systems began to integrate, what started out as a functional behavioral component of the taste aversion system, the facial movements that accompany retching, took on another

function, namely that of signaling the presence of parasites and infectious disease to others. The need for perspicuous signals added to an already substantial a set of mutually reinforcing selective pressures driving the two systems towards integration.[15]

This collection of preexisting architectural features, functional overlaps, novel selection pressures, and the need for a distinctive form of signal, all support the Entanglement thesis. Indeed, the conjunction of circumstances makes a powerful case that the human parasite avoidance mechanism gradually combined with the taste aversion mechanism, dragging their concomitant response features together in the wake.

2.3.2 Entanglement and Human Uniqueness

This positive story about how and why the two mechanisms became entangled also sheds light on why it is difficult to find anything fitting the description of human disgust in other animals: in the respects relevant to the Entanglement thesis, other species went down *different evolutionary pathways* than the one humans traversed. For instance, our closest living relatives in the animal kingdom, other primates, are omnivorous like us. Primates remain mostly reliant on foraging for sustenance, however. The evolutionary account sketched above suggests that since other primates never made the shift to hunting, scavenging and a diet high in meat, they were also never exposed to the new wave of parasites and corresponding selection pressures that would accompany such a diet. Though other primates faced the omnivores dilemma on the one hand, and the ubiquitous threat of pathogens, on the other, those sets of adaptive problems never came to coincide to the degree they did for early humans. Thus, the mechanisms designed to address those distinct problems were never forced to integrate into anything akin to the

[15] For more on information sharing, recognition and expression see Griffiths (2001, 2003), Pinker (1997, chapter 5) and Frank (1988). We will go into these issues in depth in the next chapter as well.

composite human disgust system. Finally, and as we will see in later chapters, the developing ultrasociality of humans made information sharing especially important in ways that had no counterpart in other species.

Similar reasoning suggests why these two mechanisms might not have combined in purely carnivorous species, either. Many such species have a long evolutionary history of obtaining food not only through hunting fresh meat, but during dry spells they often fall back on the option of scavenging. As a result, they would have long been endowed by natural selection with a much more durable gastro-intestinal system than our own. Being designed and conditioned to process scavenged food, such species' would be equipped to digest and deal with the types of parasites commonly found around death and decay. Humans, on the other hand, have a gastro-intestinal system originally designed for foraged foods, or at least not accustomed to rotting meat or scavenging. Thus, humans would need to avoid the sorts of parasites that actual scavengers could simply consume, and count on their gastro-intestinal system to eliminate.[16]

2.4 Conclusion: Solving the Puzzle

We began this chapter with a puzzle that arose from a tension between two views about the status of disgust when viewed from the perspective of comparative psychology. One of these, the Simple Continuity View, held that there were clear homologies to disgust in primates and other species. The other, Terror Management Theory, held that disgust is a uniquely human emotion with no counterpart in other animals.

With the Entanglement thesis, we advanced a solution to the puzzle, which shows that while both views contain a kernel of truth, each is potentially misleading and should

[16] I am thankful for a few conversations with Jonathan Haidt that helped clarify my thought on this score.

be rejected. Consider the two comparative questions posed earlier. First, are there homologies of disgust in primates and other animals, as supposed by the Simple Continuity View? According to the Entanglement thesis, the answer to this question is, indeed, yes. For, there are capacities dedicated to monitoring food intake, including mechanisms for acquiring taste aversions, found in other species. Moreover, there are also capacities devoted to protecting against parasites and disease in other species. However, the Simple Continuity View must be rejected because it is too simple: the core mechanisms involved in disgust production remain separate and distinct from each other in other species. Thus, in no other species do we find a capacity that fits the full description of disgust, containing of such diverse elements as sensitivity to contamination potency, nausea, gaping, and avoidance and sense of offensiveness.

This leads us to the second comparative question: is the emotion of disgust unique to human beings, as supposed by Terror Management Theory? According to the Entanglement thesis, the answer to this question is, again, yes. For, only in humans did these two mechanisms become entangled to form this particular emotion. However, those who subscribe to Terror Management Theory see disgust as a specifically mouth-based rejection system. As we have seen this is only half the story, and given the fascinating contamination effects that can be traced to the disease avoidance mechanism, the least interesting half in my opinion. Moreover, proponents of Terror Management Theory also see a major role for disgust in helping to manage terror, which it does by helping to repress thoughts of our animal nature and mortality that would disrupt our normal, fitness-enhancing behavior. As we have seen, though, in producing aversion to things like blood, feces, and organic decay, the emotion is actually protecting us from the

parasites and other microbes that are likely to be present there. Thus, we may conclude that the Entanglement thesis provides a better explanation than either the Simple Continuity View or Terror Management Theory of the core features of disgust, in particular the cluster of element that make up the disgust response. It also supports the controversial claim that disgust is uniquely human, but provides new grounds for this claim.

In terms of the larger goal of constructing a theory of disgust, the Entanglement thesis represents another important part of that theory, namely an ultimate explanation of the execution mechanisms that underlie production of the emotion. Like the psychological model, our evolutionary story also contributes to the goal of satisfying our pair of constraints. Recall what they are:

> *The Unity of the Response*: The characteristic disgust *response* is comprised of a number of distinct features. These features form a homeostatic cluster: they occur together as a package, and regardless of what triggers disgust on any particular occasion, once it is triggered the production of one element of the cluster is regularly accompanied by the production of the others. What accounts for the clustering of this idiosyncratic set of features? Why have these particular cognitive, behavioral and physiological elements merged into a single, unified, and apparently universally human, response type?
>
> *The Diversity of Elicitors*: A wide and surprisingly diverse range of *elicitors* trigger disgust, ranging along one dimension from the very concrete to the very abstract, along another from the universal to the culturally and individually specific, and along another from the brutely physical and inert to the highly social and interpersonal. What accounts for the pairing of such a large variety of triggering conditions to this one specific type of response?

Our ultimate explanation directly addresses the Unity of the Response desideratum. It shows that the different components of the disgust response, though they can be traced to distinct underlying mechanisms, form a homeostatic cluster because those underlying mechanisms were driven together by natural selection until they became entangled.

Together with the insight provided by the psychological model, the Unity of the Response constraint is now satisfied.

The Entanglement thesis begins addressing the Diversity of the Elicitors desideratum in a slightly more subtle way. First and foremost, it explains why the most reliable, pan-cultural indicators of disease make good candidates to be innately specified universal elicitors of disgust. It also explains the prominence of food in the elicitor pool. Most interestingly, the entanglement of these two mechanisms created a psychological system that was optimally positioned to accrue novel functions as humans became increasingly social creatures. That system was able to reliably produce a specific, aversive pattern of behavior, and that is just the sort of thing that a tinkering Mother Nature is liable to exploit and build upon. The system was also equipped with the beginnings of a flexible elicitor system. The disease and parasite avoidance mechanism was antecedently sensitive to a wide range of cues that might indicate potential for infection, having to do with places, substances, and phenotypic abnormalities in others. This is in contrast with the restricted set of conditions associated with the food rejection and acquired taste aversion mechanism, for instance. Finally, once the gape face was co-opted to serve as a signal, a nascent information sharing system was in place. Again, this is just the sort of feature that Mother Nature, in her willingness to use whatever is available to serve her purposes, is liable to build on when new purposes arise. It is by reference to these novel purposes that many of the other elicitors of disgust, especially those having to do with social norms, moral judgment and ethnocentrism, are to be explained.

Completely satisfying the Diversity of the Elicitors constraint will require us to look more closely at how, exactly, Mother Nature exploited this new response and the flexibility of its elicitor system. More specifically, making sense of the elicitor pool will require that we consider the purposes and novel functions disgust accrued. We turn to this in the next two chapters.

Chapter 3: Disagreement Over Disgustingness: Variation by way of Acquisition

3.1 Introduction

Arguments over whether or not something is disgusting can be both impassioned and difficult to resolve. A vegetarian might find a bloody rare filet mignon repulsive, while a connoisseur of fine steaks could just as easily be disgusted by tofu in all its various forms. Likewise, entire groups of people might be disgusted not just by the distinctive cuisine of other groups, but also by the very social practices and values that those groups promote. Rather than attempt to settle the disputes that might arise over such disagreements, or provide a semantic account that makes explicit the content of the claims and arguments that make them possible, we will take a different angle in what follows. Our focus in this chapter, rather, will be on the issue of acquisition.

Of course it takes a bit of a leap to get from this type of disagreement to issues about acquisition, as the connection between the two may not be obvious at first. The next section, therefore, will serve to make the relationship between the two explicit. It is devoted to motivating the general topic of acquisition, showing how the cognitive model helps frame questions about acquisition, and reviewing how the data contained in the behavioral profile bear on it. The next two sections give more detailed descriptions of some of the cognitive mechanisms that are likely to underlie the acquisition of disgust elicitors, with section 3 focusing on a few of the most well-understood mechanisms of individual acquisition. In section 4, we turn to the mechanisms that are responsible for the social acquisition and transmission of disgust elicitors. While little has been written

about these, there is a large literature on the ability to recognize expressions of emotions such as disgust. After surveying the major findings of this literature, I argue for the Empathic Acquisition thesis, which holds that the relevant mechanisms do more than just enable expression and empathic recognition, but also provide a powerful route for the transmission of cultural information and the social acquisition of disgust elicitors.

3.2 Preliminaries: Variation, Acquisition and the Cognitive Model

One of the constraints we imposed on our theory of disgust had to do with the diversity of elicitors:

> *The Diversity of Elicitors*: A wide and surprisingly diverse range of *elicitors* trigger disgust, ranging along one dimension from the very concrete to the very abstract, along another from the universal to the culturally and individually specific, and along another from the brutely physical and inert to the highly social and interpersonal. What accounts for the pairing of such a large variety of triggering conditions to this one specific type of response?

The material covered in this chapter will help in satisfying this constraint, but will not suffice to completely meet it. In what follows, we will be speaking mainly to the issue of variation, rather than diversity in general. That is, we will focus on the dimension that ranges from the universal to the culturally and individually specific. To see the difference, consider two people who are disgusted by exactly the same set of things, from the smell of putrid feces to fried tofu, to (what they consider to be) decadent liberal views on sexual mores, to Democratic political practices. There is still great diversity in those things that elicit disgust in both people. One still might wonder, about both people, why the single emotional response is elicited by such diverse types of things, that seem to have little else in common. However, there is no *variation* between the two people that requires explanation. Now imagine a third person, who loves tofu but is disgusted by beef, (what she considers to be) barbaric and oppressive conservative views on sexual

mores, and Republican political practices. This is still an impressively diverse range of things to be disgusted by, but there is also significant variation between what disgusts this third person and what disgusts the first two. Of course, explanations of diversity and variation could very well overlap – the same cognitive mechanisms could be operative in producing both phenomena. I merely distinguish between the two in order to indicate what will be emphasized below.

Now let us turn to the cognitive model presented at the end of chapter 1. Part of the utility of that model is that it allows us to pose questions about the emotion of disgust with a high degree of specificity. This includes questions about disagreement and variation. The model also allows us to frame such questions in a form that is not only more tractable, but also more focused on their psychological dimension (rather than on who is right in the relevant disagreements, for instance.) How did the vegetarian come to be disgusted by rare steaks? How did the steak eater come to be disgusted by tofu? In the picturesque language made vivid by the model, one might ask, of any particular person and any specific elicitor, how *that* elicitor got into *that* person's disgust box. Was it there innately, so that she didn't have to learn to be disgusted by, say, the smell of organic decay? Did her individual experiences, perhaps a visit to a slaughterhouse in her youth, lead her to become disgusted by beef? If she is disgusted by the Republican positions on homosexuality and abortion, is this because she acquired those elicitors from her liberal, Democratic parents and peers?

The data compiled in the behavioral profile that motivated the cognitive model suggested that some things are universally disgusting, for instance, a variety of bodily fluids, corpses, and reliable indicators of infection. The account of the evolutionary

history and primary functions of mechanisms in the disgust execution subsystem, encapsulated in the Entanglement thesis and defended in the pervious chapter, provides further reason to think that many of these universal elicitors of disgust are innately specified, part of the species typical psychological endowment of modern humans.

On the other hand, it is also a truism that different people are disgusted by different things, to the point where arguments and disagreement over whether something is *really* disgusting can become quite heated. Indeed, the behavioral profile began filling in the details of this picture, indicating that such variation occurs at both the individual and cultural level. These patterns of variation licensed the positing of mechanisms of both individual and social acquisition. The guiding inference was that variation of this sort suggests the operation of specialized learning processes that are sensitive to particular environmental cues, social or otherwise.[17] Patterns in the population level distribution of disgust elicitors are to be explained by appeal to the operation of mechanisms of social learning that support those patterns, namely by allowing the transmission and acquisition of information about what to be disgusted by. Instances of individual level variation are to be explained by appeal to differences in personal histories and the operation of mechanisms of individual learning, which acquire elicitors via direct experience and interaction with types of items, substances, and so forth. The functional role of such mechanisms was represented by their place in our cognitive model. It will be useful to recall that model:

---

[17] This is the mirror image of the (defeasible) inference that the universality of some feature suggests that it is innately specified: the presence of variation suggests the presence of capacities dedicated to learning and acquisition, and cognitive mechanisms underlying those capacities. It is also worth noting that just as universality does not always indicate innateness, not all variation suggests learning, either; differences in height or hair color indicate genetic differences, rather than variation acquired from different social environments or personal history.

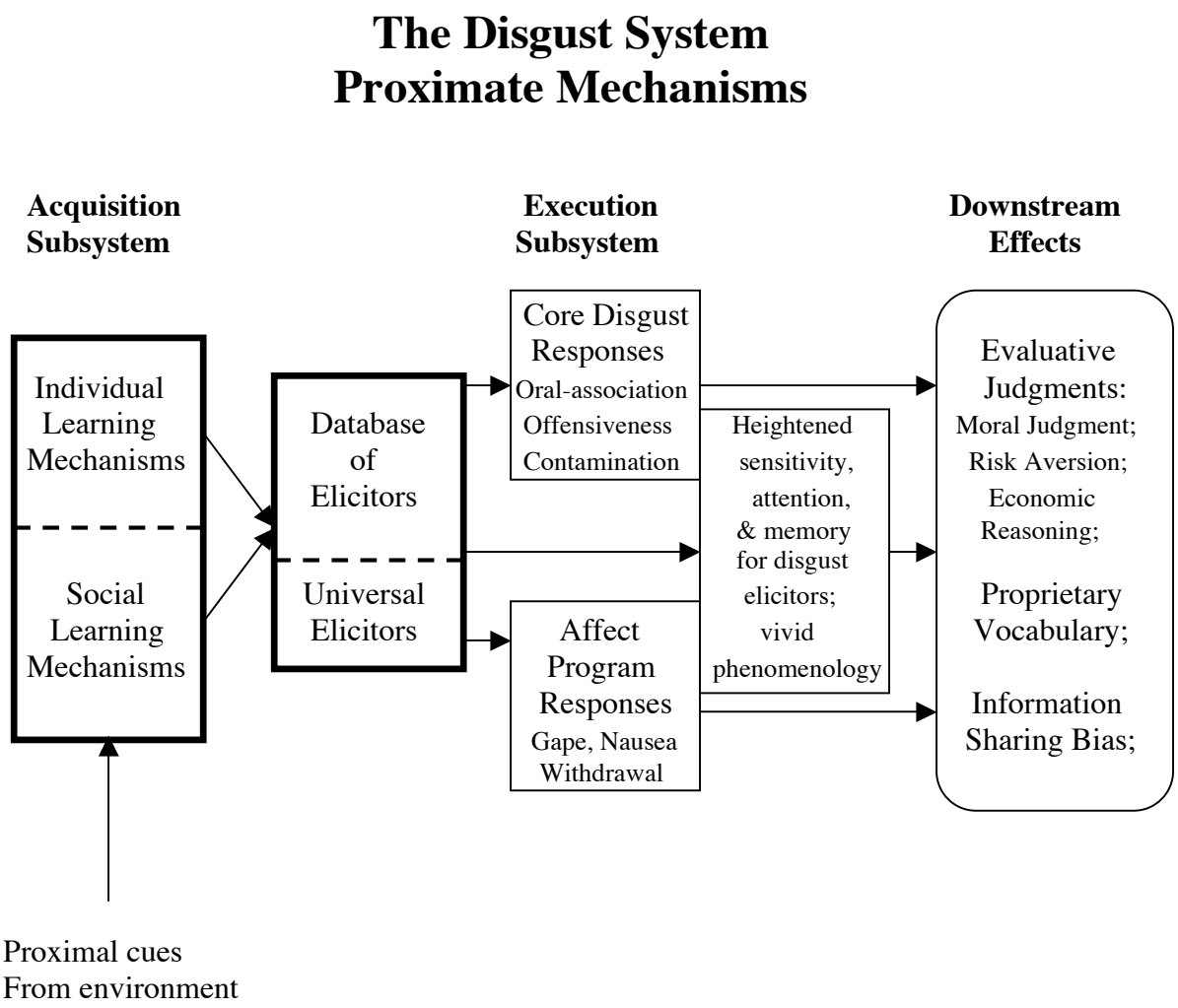


Figure 3.1
A functional level model of interlocking mechanisms that comprise the human disgust system. The arrows represent causal links between the various mechanisms.

This cognitive model held a place for a number of distinct, independently operating cognitive mechanisms responsible for acquiring disgust elicitors. Generally speaking, the function of both individual and social acquisition mechanisms is to pick up on relevant cues and patterns from the surrounding environment, social or otherwise, and infer from them new contents for the disgust database. The division between individual and social acquisition is based on the proximal cues to which a particular mechanism is sensitive: the former involve acquisition of elicitors unmediated by social transmission or

other people. Individual and social learning probably represent poles along a continuum rather than anything categorical, but the distinction is clear enough to be useful for organizational purposes.

As it was initially presented, the acquisition subsystem allowed for a plurality of acquisition routes and mechanisms, but remained agnostic as to their number and individual character. The main goal of this chapter is to describe some of the particular mechanisms with a greater degree of specificity.

3.3 Individual Learning Acquisition

It appears that a number of different mechanisms can serve to deliver a new elicitor to the disgust database. Indeed, it appears that much of the flexibility of acquisition itself likes in this plurality of mechanisms. Specifically, routes of acquisition that comfortably fit under the heading of individual learning are far from systematic, and are thus likely subserved by a number of heterogeneous mechanisms.

One specific mechanism of individual learning was mentioned in chapter 2, and evolved in tandem with the taste aversions they produce. While this mechanism itself is innate, it provides one route, albeit a highly restricted one, by which organisms can acquire aversions to foods that have caused gastro-intestinal stress.[18] Recall that acquired taste aversions provide a way to narrow down culinary options by implementing a "once bitten, twice shy" rule – a type of food that has induced sickness in the past is avoided in the future. The learning mechanisms associated with acquired taste aversions underlie a type of "one shot learning", that is, they require only a single trial to acquire an aversion to a new type of food. If consumption of a particular food is accompanied by gastro-

[18] As mentioned previously, it remains an open question of whether acquired taste aversions produce disgust straightaway, or if they make it more likely to become fully disgusted by the culprit type of food.

intestinal stress, even as far as 12 hours after consumption, an aversion to that food is developed.

Other routes of acquisition are less restricted. Processes as simple as classical and operant conditioning could lead to the acquisition of a new disgust elicitor. Rozin points out that intense experiences can lead individuals to become disgusted by substances that did not disgust them previously. He uses the example of a visit to a slaughterhouse, which might be so gruesome that one is forever after repulsed by beef. He also speculates that new elicitors can be acquired more circuitously, as when one is convinced by rational argumentation of the immorality or disgustingness of a practice such as smoking or eating meat (Rozin 1997). Fessler et al. (2003) lend some indirect empirical support to this speculation when they conclude, based on a web-based, self-report survey of nearly a thousand adults, that "'moral vegetarians' disgust reactions to meat are caused by, rather than the cause of, their moral beliefs." In other words, in many cases of moral vegetarianism, propositional reasoning is instrumental in the acquisition of meat as a new disgust elicitor.

Another route of individual acquisition appears to involve a kin recognition mechanism. In cases of this sort, the kin recognition system operates in conjunction with the disgust system to help prevent incest. The adaptive problem here is familiar: inbreeding leads to a decrease in genetic diversity and allows for the expression of recessive genes, which in turn diminishes the health and fitness of inbred offspring. Westermarck (1891) originally posited that disgust played an important role in human incest avoidance, and subsequent research has, in broad strokes, vindicated his

conjecture. Experimental data establish the hypothesis that the emotion of disgust plays some crucial role (Lieberman et al. 2002, Fessler & Navarrette 2004).

Furthermore, some unusual marriage arrangements from various cultures serve as natural experiments bearing on this issue and provide startling insight into the role played by this kin recognition mechanism. Marriage rates between boys and girls brought up in close personal proximity with one another, i.e. children unrelated but raised as if they were siblings, are inordinately low. In one case of this, Shepher (1983) shows that marriage amongst Israeli kibbutz age mates is extremely rare. In a similar case, Wolf (1970) shows that Taiwanese minor marriages were extremely unsuccessful. In these cases, marriages were arranged while the eventual bride and groom were still young. Moreover, once the pairing was arranged, the family of the eventual groom would adopt the eventual bride. This usually occurred between a few months to three years of age, and the two were to be married once they reached the appropriate age. However, as mentioned above, the success rates of these marriages were extremely low. These examples are taken to show the kin recognition mechanism misfiring and giving false positives, but in so doing they illustrate the principles that govern it. The mechanism is automatically calibrated during an innately specified developmental window early in life. In the case of incest avoidance, that calibration is accompanied by the acquisition of an individually learned disgust elicitor. The kin mechanism recruits disgust to block potential sexual attraction between crib mates later in life.

More speculatively, the phenomenon of genetic sexual attraction provides examples of what happens when the mechanism is not properly calibrated and gives false negatives. In cases of genetic sexual attraction, opposite sex siblings separated at birth

and raised apart, who then meet up later in life, can find themselves dealing with unwelcome but strong feelings of sexual attraction to each other. A similar phenomenon has also occurred between parents and children given up to adoption, when they were reunited later in life. Though no systematic study has been conducted on these phenomena, it is becoming more acknowledged and widely addressed by the institutions that regulate adoption.[19] The Westermarck view suggests an explanatory hypothesis: what these cases have in common is that kin recognition mechanisms are not properly calibrated. As a result, disgust is not recruited to muffle sexual attraction as it usually does, and that attraction is being allowed to express itself in these anomalous cases.

This list of individual learning acquisition mechanisms is not meant to be exhaustive. Other routes and mechanisms are likely to be uncovered with further research. One point to take away, however, is how ad hoc and unsystematic the collection is. In this respect, it also serves to illustrate how, like the execution system, even the acquisition subsystem of disgust bears the marks of a kludge, and how it has been co-opted to carry out functions it did not initially evolve to perform.

3.4 Social Learning Acquisition

By contrast with individual learning, new elicitors may also be learned from conspecifics or acquired via some form of social transmission. Along with familiar anecdotes, preliminary data confirm that at the population level, many disgust elicitors exhibit a patterned variation common to other types of culturally transmitted items: uniformity within cultures but diversity across cultures (for instance Haidt et al. 1993, Haidt et al. 1997, Haidt and Hersh 2001, Rozin and Segal 2003; see also Miller 1997,

[19] For more details see: www.reunite.com/adoption-records/genetic-sexual-attraction.html.

Elias 1939). Such patterns are striking, and neither individual learning nor appeal to an innate endowment alone can plausibly account for them. Rather, the patterns suggest that cultural transmission effects the population level distribution of disgust elicitors, and thus at the individual level, social learning plays an important role in their acquisition.

Mechanisms of acquisition that comfortably fit under the heading of social learning appear not to be as multifarious or piecemeal as those underlying individual learning, but the mechanisms (and more than likely the selection pressures that shaped them) are much more complex. Above we noted that the disgust execution subsystem has features that made it a prominent candidate for being co-opted and put to new purposes. Novel selection pressures took advantage of this susceptibility, and conferred many new functions on the disgust system. As we will see in the next chapter, these selection pressures were largely generated by early hominids' increasing reliance on social interactions with conspecifics, and the complex social structures created by that reliance. At the simplest and most general level, sociality requires some degree of communication and information transmission between conspecifics. As a result, mechanisms that facilitate these activities were selected for, and refined to better perform such functions.

In what follows, we will look at more detailed proposals about the more rudimentary types of proximate mechanisms that could have allowed for this communication and information transmission. Paradoxically, little if any work has been explicitly devoted to the issue of social acquisition. However, much of what we'll cover has recently been used elsewhere in debates about the types of mechanisms that might underlie our "theory of mind" or "mentalizing" capacities (see Goldman 2006, Nichols &

Stich 2004). What will be distinctive of the treatment here is that I argue these mechanisms of expression and empathic recognition also provide an especially powerful route for the social *acquisition* of disgust elicitors.

3.4.1 Emotional Expression and Facial Recognition

Humans are extremely social creatures. This sociality lies near the heart of human nature, and so informs aspects of our emotional and cognitive makeup far beyond disgust, or any other single component of our psychology. Natural selection has equipped us with a variety of mechanisms with which to extract information about the social environment from the behavior of our conspecifics, as well as mechanisms that serve to relay information about ourselves to those same conspecifics. Many of these mechanisms are sophisticated, highly specialized, and uniquely human. Examples of these include some of the central mechanisms involved in mindreading, and many of the mechanisms underlying language use and acquisition. While human language is easily the richest and most complicated form of social signaling found in nature, simpler capacities for nonverbal communication are ubiquitous. Body language might be the most obvious and intuitive of these. Many of the nonverbal capacities found in humans are also shared with other primates, in whom homologous versions can be found in slightly different, often cruder form. For example, the attention of other organisms is often a good source of information about their intentions and immanent behavior. Mechanisms underlying gaze monitoring track the visual attention of other organisms, thus gaining useful information about the immediate social environment. Such mechanisms can be found in many mammals, primates and humans included (for instance, see Tomasello 2001).

Another important class of mechanisms that can be found in both humans and other primates concern emotional expression and recognition. In expressing an emotion, the "sender" organism signals information about its own internal state to those "receivers" that recognize the expression. The mechanisms underlying expression and recognition have been extensively studied in humans (Ekman 1992, 2003), and have also been investigated in other primates such as rhesus monkeys (Mason 1985, see also Griffiths 2003). Their significance was first remarked upon by Darwin himself (1872). Darwin speculated, and later research has largely confirmed, that in humans there are distinct, universal, and universally recognizable facial expressions associated with a core set of "basic" emotions. Disgust has long been recognized as a member of this set. Indeed, Rozin et al. (2000) claim that it appears on nearly every list of basic emotions that has at least four members, from Darwin's own list onward. As such, we should expect to find mechanisms underlying expression and recognition in the human disgust system.

3.4.2.1 Expression: Sending Signals

And of course, we do. Many of those mechanisms responsible for expression are located in the execution subsystem. Like other emotions, disgust is expressed by the behavior and physical symptoms that reliably accompany it. Many characteristic features of the behavioral response not only fulfill a direct adaptive purpose, but also serve to signal others that the disgust system has been activated. While it is unlikely that these features were originally produced to convey information to others, the fact that they are

highly correlated with the activation of the disgust system lead them to be easily co-opted to serve as signals.[20]

In the case of humans, the face carries an inordinate amount of information. As we might expect, then, the most obvious and recognizable expression of disgust is the gape face. As seen above, the gape was originally a component of the taste aversion subsystem, where it was a precursor to vomiting, an adaptive response to potentially toxic food. Since it is more easily identifiable than the other elements of the behavioral response that it regularly occurs with, such as offensiveness, withdrawal, contamination sensitivity, etc., the gape face came to also serve as an effective signal of their occurrence. It thus doubles as a reliable source of information for receivers about the internal state and immanent behavior of senders, as well as a signal for potentially infectious or toxic substances in the immediate environment.

In this respect, the gape face is akin to many other traits and behaviors found throughout nature that did not originate to serve as signals, but later came to perform important signaling functions. Such "passive" signals are contrasted with "active" signals, the later of which were selected for the express purpose of conveying information (Frank 1988). For instance, an oft-cited example of a passive signal is the depth of a toad's croak. For brute reasons concerning acoustics and the physics of sound, the depth of a male toad's croak is correlated with his size. Croaks were originally used simply to attract nearby female toads to mate. However, since the depth of a croak also carried information relevant to mating and fitness, namely about the size of the male making it, females began to discriminate, and prefer the deeper croaks of those larger toads. Thus

[20] See Hinde 1985a and 1985b for early discussions of the link between expression and signaling.

croak depth, initially a mere byproduct, accrued the function of being an important signal in the mating dynamics of toads.

Passive signals have another feature that often distinguishes them from active signals: they are costly and difficult to fake. Again, the depth of a toad croak is physically constrained by the size of the cavity and vocal chords that produce it. This makes it impossible for a small toad to fake being a large toad with a "false" signal, i.e. by producing a deeper croak than a larger one. Surprisingly, there is evidence that the gape face shares this feature of passive signals as well. Like many other characteristic faces of the basic emotions, the gape appears to be subject to facial feedback. Exaggerating or suppressing a facial expression can enhance or diminish the strength of an emotional response via a feedback mechanism between the facial component and other coordinated components of the execution subsystem (Laird and Bressler 1992, Hatfield et al. 1994). Moreover, voluntarily making an emotion face, a single component of the response, can serve to activate the execution mechanism and initiate the entire coordinated response (Coan and Allen 2003; Levenson 1992 calls this plasticity of elicitation). In other words, acting like you are experiencing a particular emotion can actually cause you to experience it; making a gape face when you are not otherwise disgusted can, via facial feedback, cause you to become disgusted. Of course, faking a gape face need not induce disgust with the same intensity as, say, a whiff of fresh manure. This is not to say gaping does not activate the disgust execution system, however. Such activation might be mild, even subthreshold and so below the level of awareness.

Even when the activation of the disgust system is subconscious, however, that emotion is outwardly expressed. One might intuitively think of these along the lines of "tells", the sort of things that very good poker players and other readers of body language are able to pick up on. They are involuntary, and their duration can be as short as 40 milliseconds, so they are obviously not very exaggerated. These, too, often occur without the conscious awareness of the sender. Indeed, they often "leak" and flash across the face despite deliberate attempts to hide them and the internal states they signal (Ekman 2003). In this respect, gape face signals are difficult to fake along another dimension as well. Since microexpressions are nearly impossible to completely suppress, it is nearly impossible to fake that you are *not* disgusted when you actually *are*.

This vestige of retching, then, appears to be fairly well fitted to the role of signaling. In fact, though the gape face is a passive signal, and thus did not originate to convey information to conspecifics, it appears to have become more refined in its role as a signaling device in other ways as well. It can be a fairly subtle source of information for receivers, not just about the internal state of the sender, but about the item in the external environment that is triggering the sender's disgust. Individual components of the gape face, such as the nose wrinkle, tongue extrusion and raised upper lip, can be variously exaggerated or deemphasized depending on whether the primary source of disgust smells bad, looks disgusting, and so forth (Rozin et al. 1994). Thus the facial expression of disgust has become more sophisticated, as idiosyncrasies in particular gape faces can convey more discriminated information about the state the sender's offending environment.

While the face is the richest and most fine-grained source of information for receivers, emotions are expressed through a variety of other behavioral channels as well. Again, these can be thought of, intuitively, as the other components of body language and other types of nonverbal communication. In addition to actions such as a quick withdrawal, other features of a sender's bearing and body language such as orientation and posture, as well as nonverbal elements of vocal expression such as intonation and cadence, can signal the occurrence of a specific emotion. With respect to disgust, these signals would be the various manifestations of offensiveness, such as leaning away from the offending item. Like the gape face, most of these features did not originate to convey information to conspecifics, but where physical symptoms and byproducts of disgust that later acquired a signaling function. Though there has not been as much research on these features, there is reason to believe that mechanisms of vocal, postural, and muscular contraction feedback make the signals they send difficult to fake for the same reasons that facial feedback makes the gape face difficult to fake. Likewise, if plasticity of elicitation is not limited to facial expressions, mimicking the body language and cadence of disgust can activate the entire suite of responses, causing the would-be faker to actually become disgusted (see Hatfield et al. 1994 for an overview).

The mechanisms of emotional expression contain one more subtlety. Expression of disgust through all of these various channels, like expression of other emotions, may be modulated not just in response to properties of the eliciting item, but also in response to important features of the social environment in which disgust is being expressed. Griffiths has recently argued that basic emotions "may be Machiavellian all the way down" (Griffiths 2003). He argues that in both humans and primates such as rhesus

monkeys, the expressive components of emotional responses are often exaggerated or suppressed in ways that are quite sensitive to socially strategic aspects of the triggering situation. Since this variability is found in other primates, he argues that it is not a function of cultural display rules or the intervention of higher-level cognitive processes. Instead, it indicates that greater sophistication than was previously thought to exist in the more rudimentary mechanisms employed by affect programs. If this line of thought is on the right track and applies to disgust, then the mechanisms underlying the expressive responses of senders can automatically calibrate particular expressions to control what and how much information is conveyed to particular receivers. Subtleties such as these provide further evidence of how the signaling functions accrued by the disgust execution system were further refined in the face of the selective pressures associated with increased sociality.

3.4.2.2 Recognition: Decoding Signals

Expression and recognition go hand in hand, of course: for some trait or behavior to effectively serve as a signal, receivers must be able to detect it, recognize it as such, and extract the relevant information from it. As Darwin saw, the gape face is not just a distinctive and universal expression of disgust, but is universally recognizable as well. Like those responsible for execution and expression, the mechanisms underlying the ability to recognize social cues of disgust have some surprisingly complicated features. For instance, these, too, appear to be innate. Since the gape face is universally recognizable, the mechanisms responsible for the capacity to recognize it as such are likely universally present in all normal humans. Emotional recognition appears to be operative, to some degree, very early in ontogeny, as children have been show to be

sensitive to the facial expressions of caregivers by the time they reach 12 months (Bandura 1992). Together these facts also suggest the core meaning of emotion faces like the gape do not have to be explicitly taught or learned, but are rather known innately.

There is also evidence indicating that there are mechanisms specifically dedicated to recognizing expressions of disgust. Much of this evidence turns on the discovery of selective impairments of one capacity but not another. For instance, facial emotional expression recognition, on the one hand, appears to be distinct from the system responsible for facial identity recognition, on the other. This is suggested by the fact that subjects with lesions to specific brain areas retain their capacity to identify individual people by their faces, but lose the ability to pick out many facial expressions of emotion (Keane et al. 2002). Recognition of disgust expressions is distinguishable from capacities for recognizing other emotions as well. Huntington's disease significantly impedes the ability to recognize the gape face as an expression of disgust, but does not have similarly strong effects for other characteristic emotion faces (Sprengelmeyer et al. 1996). Obsessive-compulsive disorder is accompanied by a similar selective impairment of the ability to recognize disgust expressions, but not other emotion faces (Sprengelmeyer et al. 1997). While the capacity to recognize expressions of disgust as such slightly increases into old age, recognition of fear and anger decreases; other emotions remain stable (Calder et al. 2001).

Disgust recognition is also distinct from other emotions in its neural substrate. While phenomena involving many other emotions have been linked to the amygdala, recognition of disgust is associated with activation of the putamen and insula (Phillips et al. 1997). Damage to these areas selectively impairs recognition of disgust, leaving the

ability to recognize other emotion expressions intact (Calder et al. 2001, Adolphs et al. 2003). Hence, disgust recognition is not only functionally specific, but and neurally specific as well.

In addition to these forms of specificity, the disgust recognition capacity is also multimodal. Like specificity, this property of the capacity has also been revealed by evidence from neural correlates and impairments. Damage to the putamen and insula impairs not just the ability to recognize gape faces as expressions of disgust, but expressions of disgust over a variety of other modalities as well. For instance, vocal expressions like the sound of someone else retching have been used in several studies. This suggests that the core mechanisms subserving recognition of disgust expressions are not only specific to disgust, but are also multimodal, and thus separate from the components of any single perceptual system (i.e. visual, auditory, tactile) that they employ (Calder et al 2000, Keane et al. 2002).

A final feature of disgust recognition is that it is often *empathic*: recognition of a sender's emotion expression is accompanied by slight production of the emotion in the receiver. In other words, when recognizing someone else's disgust, you often come to feel disgusted yourself. Disgust does not appear unique in this respect; recognition of other emotions is empathic as well, i.e. recognizing anger is accompanied by the production of anger, recognizing fear is accompanied by the production of fear, and so forth (Hatfield et al. 1994, also see Goldman 2006 for a lucid overview). In addition to anecdote and introspection, some of the most impressive evidence of empathic recognition comes from fMRI studies. Indeed, recent studies confirm that the same brain

areas, the putamen and especially the insula, are active in both the experience of disgust and the recognition of another's disgust (Wicker et al. 2003).

It is quite striking that these same neural areas play a role in both generating feelings of disgust and recognizing expressions of disgust in others. Moreover, this points towards a single hypothesis that explains all of the features of the disgust recognition capacity mentioned so far, its functional and neural specificity, its multimodality, and the fact that it is empathic: many of the core mechanisms subserving disgust recognition are the same mechanisms found in the execution subsystem. The recognition subsystem appears to exploit the disgust execution subsystem in some crucial respect.

This explanation once again turns on mechanisms becoming multifunctional as they were co-opted and put to new uses. Some of the core mechanisms dedicated to execution and core mechanisms dedicated to recognition appear to be one and the same. Indeed, even the neural realization of the mechanisms appears to be identical. Those mechanisms accrued the new functions associated with recognition after they had already been shaped by the selection forces responsible for the disgust execution system. Thus the multimodality of disgust recognition is derived from the multimodality of the mechanisms underlying disgust execution: the execution system is distinct from the different perceptual systems, but can use all of them to detect the presence of disgust elicitors. Disgust recognition is functionally specific because it requires the disgust execution subsystem, rather than, say, the anger or fear subsystem; disgust recognition can fail while anger and fear recognition remain unimpeded, and disgust recognition can succeed while other forms of recognition fail or fall off.

There is support for this hypothesis beyond its ability to explain the other features of disgust recognition. The brain areas implicated in recognition, the putamen and especially the insula, had been previously identified on independent grounds with gustatory and olfactory functions, nausea and taste aversion (Small et al. 1999). One of their important functions has been identified as translating unpleasant sensory input into visceromotor reactions and the feelings of unpleasantness and revulsion that accompany disgust. Interestingly, taste aversions have been linked to multiple modalities as well (Bernstein 1999). Perhaps most convincing are paired deficits between recognition and disgust production. Failure in the ability to recognize gapes as expressions of disgust accompanies failure in the ability to produce disgust itself (see Goldman & Sripada 2005). All of this strengthens the case for the hypothesis that disgust recognition crucially involves activation of the disgust execution system.

The exact role of the execution system is still unresolved, however. At least two kinds of hypotheses about the causal sequence involved in empathic recognition are clearly discernible (see Goldman & Sripada 2005 for graphic representations of some models and a more detailed discussion of the space of theoretic options). According to the first, a sender first recognizes that a sender is experiencing a particular emotion. Once the receiver makes a cognitive match between expression and emotion of a sender, the execution subsystem of the relevant emotion is then activated, and the recognition thereby becomes empathic. Thus, an initial 'cold' recognition of a sender as disgusted then activates the receiver's execution subsystem as something of a downstream afterthought, causing the receiver to also become disgusted.

According to a second kind of hypothesis, a receiver's feeling disgusted is a crucial causal antecedent in recognizing, for instance, a gape face as an expression of disgust. Hatfield et al. (1994) endorse a hypothesis of this form for recognition of a variety of emotions. In addition to marshalling an impressive amount of evidence for the phenomenon of empathic recognition, they further endorse an explanation that places imitation or mimicry and the mechanisms of facial feedback at the beginning of the causal sequence, giving rise to emphatic recognition. On this account, the first step in empathic recognition on the part of the receiver is to behaviorally mimic the expression of the sender (see Walcott 1991 for evidence that motor mimicry plays an important role in facial recognition). This can occur via a number of channels, facial, postural, etc., and while it is often subthreshold – receivers might mimic facial expressions of senders with their own microexpressions, for instance – it often occurs continuously and automatically. Next in the causal sequence, once receiver and sender are in sync, the receiver comes to experience the emotion being mimicked via the relevant feedback mechanisms, facial, postural, and so forth. Finally, in part due to his own internal state, the receiver is able to recognize the emotion being expressed by the sender. Making a match between behavioral cues and particular emotions is facilitated by the fact that the receiver is put in a similar affect state, and is even, to some degree, experiencing the emotion as well as perceiving it in another. In terms of causal sequence, activation of the execution system precedes the full recognition of disgust expressions as such.

Perhaps even more interestingly on this account, full-blown cognitive recognition of an expression need not occur for a sender to 'infect' a receiver with his emotional state. The entire process of expression, mimicry, and feedback may take place beneath

the level of conscious awareness. Appropriately, Hatfield et al. (1994) avoid overusing the term "recognition", which carries connotations of sophisticated conscious achievement in its vernacular usage, and instead dub the phenomenon "emotional contagion".

3.5 The Empathic Acquisition Thesis

Whatever the activation sequence of the mechanisms underlying recognition of disgust expressions – and nothing bars each type of hypothesis being correct on different occasions, or other occasions of recognition not being empathic at all – the fact that recognition is often empathic has implications for the social acquisition of disgust elicitors. But first, we should note that the discussion thus far has not been cast explicitly in terms of acquisition, but expression and recognition. In advancing the Empathic Acquisition thesis, I am suggesting that the mechanisms underlying expression and empathic recognition are the same mechanisms that allow for the both the transmission and acquisition of disgust elicitors.

It is not such a large step from a receiver's recognizing the disgust felt by others, to that receiver's acquiring a disposition to become directly disgusted by whatever primary source of disgust is offending the sender. While other routes of social acquisition are of course available, for instance via purely verbal routes, these mechanisms of expression and recognition can provide the most important and rudimentary route for socially acquiring new disgust elicitors.

On this general picture, acquisition may often be a matter of cutting out the middleman, eliminating the mediating role of the sender or cultural parent in inducing disgust towards some item. When a receiver acquires a new primary elicitor in this way,

he gains the disposition to become directly disgusted by a type of item, even in the absence of the cultural parent. This form of acquisition could occur via the conjunction of recurring episodes of empathic recognition and some unembellished form of conditioning, as the feeling of disgust is repeatedly paired with the primary elicitor. Alternatively, it could be more complicated, going by way of currently undiscovered means. As we shall see below, acquiring a new primary elicitor from others may have come to be assisted by cognitive biases of various sorts as well. Indeed, one has already been discussed. Rozin et al. (1994) implies that along with a variety of other social cues, the structure of particular gape faces themselves can yield information about the nature of the primary elicitor a particular case. This would help direct receiver's attention to the relevant item in the environment, and would thus facilitate the acquisition of that particular item as a primary elicitor.

Upon reflection, we can see that many features of expression and recognition would make this a powerful channel for transmitting and acquiring elicitors. The expressive signals themselves are difficult to fake, since body language and microexpressions belie what would be false negatives, while facial and other forms of feedback initiate genuine feelings of disgust in what would otherwise be false positives. Empathic recognition provides a powerful form of acquisition for a number of reasons. Extracting information from expressions not only involves a receiver detecting social cues of disgust, but entering into a genuine affective state that is type similar to the affective state of the sender or cultural parent. Entering such an affective state is likely to have direct and immediate effects on the sender's cognitive and motivational makeup. Moreover, we have seen independent evidence that much of this process occurs

automatically, with many of the mechanisms involved operating without conscious attention or effort. In this, elicitor acquisition is able to bypass any stage of explicit inference, and thus many of the more stringent epistemic norms that usually regulate such inferences.

Many might balk at the idea that something could or should come to be considered disgusting simply because others think it so, especially once this inference type, or particular instances of it, are made explicit and subjected to critical scrutiny. By and large the particular instances are not, though. Since, on the Empathic Acquisition thesis, mechanisms operating outside of conscious awareness subserve the acquisition process, elicitors can enter the disgust database without going through any rational checkpoint. Since recognition is empathic, senders can have strong and immediate effects on receivers and their mental states. Together, these two features of acquisition can lead to social pressures having a usually strong influence in shaping population level distribution of disgust elicitors.[21]

3.6 Conclusion

The aim of this chapter was to describe in more detail some of the specific mechanisms that underlie capacities to acquire disgust elicitors. We used the cognitive model to motivate a more psychologically oriented perspective on issues like disagreement, variation, and acquisition, and showed how an account of the mechanisms of individual and social acquisition can begin to satisfy the Diversity of Elicitors constraint we accepted on our overall theory of disgust. We have also seen that a

[21] Of course, social pressures alone do not determine, unopposed, the population level distribution of disgust elicitors. Many are universally present because they are innately specified. Moreover, cognitive features other than explicitly held epistemic norms, for instance context and content biases, can and most likely do influence the distribution and dynamics of disgust elicitors.

plurality of different mechanisms have been linked to disgust acquisition, and that their number and character further illustrate the kludge-like nature of the emotion in general. Finally, in defending the Empathic Acquisition thesis, we have argued that much exciting research on the humanly universal capacities to express and recognition emotions can be brought to bear on the issue of acquisition of disgust elicitors, and other related issues, including the social dynamics that disgust might influence.

We can express the progress we have made by updating our cognitive model:

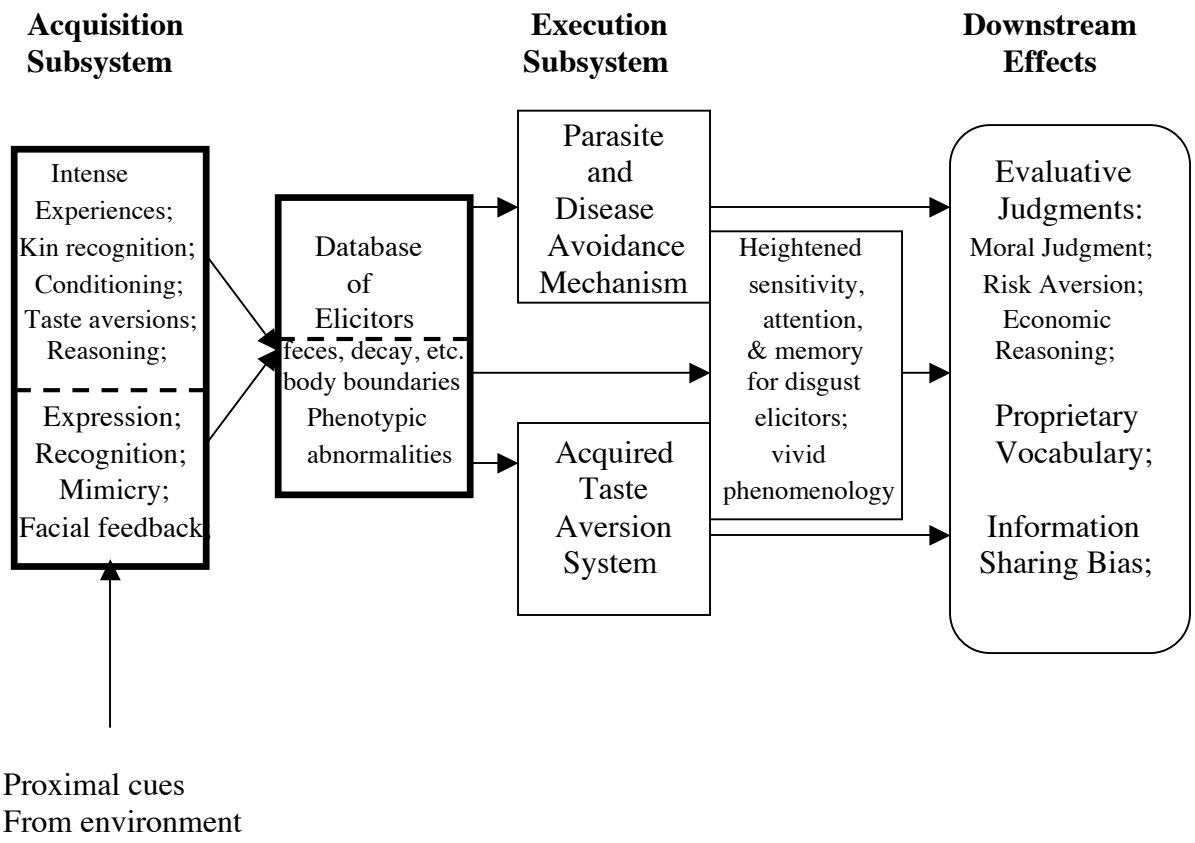


Figure 3.2
A functional level model of interlocking mechanisms that comprise the human disgust system. The arrows represent causal links between the various mechanisms.

Let us conclude with some more speculative thoughts, then. Social pressures transmitted via the mechanisms of disgust recognition and expression can also have recognizable effects on the attitudes towards, and thus dynamics surrounding, behaviors that become elicitors of disgust. For instance, Rozin (1997) points out that in many cases of moralization, disgust has played an important role in altering attitudes towards practices such as smoking and eating meat. Under the rubric of "the looping effects of human kinds," the philosopher Ian Hacking has described a set of social dynamics that

emerges from the very act of classifying different types of people. He speculates that such social dynamics, which take the form of a feedback loop between classification and those so classified, are accelerated and amplified when the larger population regards members of a kind as deviant and worthy of condemnation. Two examples of such human kinds discussed by Hacking are child abusers and homosexuals. We could easily include smokers on this list, and note that in all three cases the emotion associated with condemnation of the practice is disgust. Though ripe with potential implications, little is systematically known about these social dynamics or the cognitive and emotional engines that drive them. This is an area of inquiry that certainly deserves more attention.

In general, the patterned variation of elicitors we started with, exhibiting within group conformity and across group diversity, suggests that people are quite sensitive to and influenced by what others in their community are disgusted by. The Empathic Acqusition thesis showed how the mechanisms of emotion expression, recognition, and contamination all have a role to play in an account of how one may come to be disgusted by the same things that disgust the other members of one's community and culture, and thus a first step in explaining how those population level patterns and dynamics might be generated and sustained.

Chapter 4: Moral Disgust and Tribal Instincts: A Byproduct Hypothesis

4.1 Introduction

Most of the philosophic interest in disgust stems from its relation to morality (McDowell 1985, 1987, D'Arms & Jacobson 2000, 2005, Knapp 2003, Nichols 2004, Nussbaum 2004). Additionally, some of the most compelling recent empirical evidence about disgust has begun to map out the way disgust influences morality via its interactions with social norms and evaluative judgments (Haidt et al. 1993, Haidt et al. 1997, Rozin et al. 1999, Nichols 2002a, 2004, Schnall et al. manuscript). These discussions, while interesting in their own right, also give rise to further questions. Most generally, what *is* the relationship between morality and disgust?

This question about the relationship between morality and disgust can be read in two ways. On the one hand, one might be curious about a cluster of *prescriptive* issues that center on the question of how and whether feelings of disgust *should* enter into our moral deliberations, how they should be weighted in our considered moral judgments, or how they should influence legal theory, the legal system, and other institutions. Normative issues like these have been debated in recent years by authors such as Martha Nussbaum (2004), Leon Kass (2002) and William Miller (1997).[22] On the other hand, one might be curious about the cluster of *descriptive* and *explanatory* issues concerning

[22] Another class of questions about disgust that arises in moral theory falls within the domain of metaethics rather than normative and applied ethics. These questions, such as what is the relationship between the disgust response and the property of disgustingness, will be addressed in later chapters.

disgust, and the roles it, as matter of fact, *does* play in areas associated with morality, as revealed by empirical work in experimental psychology and cultural anthropology. These include intuitive evaluations and the regulation of a range of social interactions.

The focus of this chapter lies firmly on the second, descriptive and explanatory cluster of issues, where many questions remain open. For instance, some researchers talk as if moral disgust was a thing apart from basic disgust (Rozin et al. 2000), but it is not yet clear how to draw a principled distinction between the two. We might ask a number of questions related to this: how might the intuitive difference between moral disgust and basic disgust be understood from a psychological point of view? How did disgust come to be involved in morality in the first place? Some researchers worry that so-called moral disgust is not actually disgust at all, and hold out the possibility that while things like feces and spoiled meat are true elicitors of the emotion, talk of disgust in conjunction with moral issues is merely metaphorical (Nabi 2002, Bloom 2004; also see Haidt et al. 1997 for discussions of related worries).

This last worry is understandable, but as we saw in previous chapters, mounting behavioral and neurological evidence points in the same direction as colloquial usage of the term 'disgust' on this score, suggesting that the extension of disgust to these other domains is more than metaphorical. Rather, the complete suite of elements comprising the disgust response is produced by a very wide range of elicitors, from feces and spoiled meat to deviant sexual practices, violations of certain social norms, and ethnic boundary markers and the outgroup members who bear them (see especially Borg et al. 2006). Concerns raised by the other questions, however, are not so easily settled. For, as we have seen from our evolutionary perspective, disgust initially evolved to protect humans

from poisons and parasites, rather than to serve any overtly moral purposes. Unlike more socially-oriented, "Machiavellian" emotions such as envy, gratitude, or love, it does not appear to have originally evolved out of the strategic push and pull of social interaction at all.[23] Nevertheless, it is evident that disgust has come to play an important and systematic role in our moral psychology. In addition to its primary functions of protecting humans from poisons and parasites, therefore, the emotion must have acquired auxiliary functions connected to moral judgment, moral norms, and so forth. Clarifying the character of these functions, the way disgust performs them, and the nature of the putatively moral elicitors constitutes the final desideratum for our theory of disgust.

Here is how we will proceed. In order to meet this desideratum, and to shed light on the specific ways disgust interacts with morality, we will place the theory of disgust as we have developed it so far within the context of gene culture coevolutionary theory. This work explores the implications of humans' reliance on culturally transmitted information and the dynamics of group living. These interconnected phenomena are fascinatingly complicated (see Richerson & Boyd 2005 for an accessible overview). More importantly, the theory holds that both of these factors helped create a unique set of conditions, wherein cultural evolutionary pressures interact with natural selection, giving rise to a blend of forces that greatly complicates the evolution of human beings, and of human psychology in particular. Answering our questions about moral disgust will require that we look closely at these sorts of considerations, and especially at one component of gene culture coevolutionary theory called the tribal instincts hypothesis, which we will explore in section 2. Once we have developed these ideas in sufficient

[23] See Frank 1988 and Pinker 1997 for more on the role of the social emotions in strategic interaction.

detail, we will turn our attention in section 3 back to disgust, and elaborate on what I will call the Co-opt thesis. We will consider how the emotion acquired its moral valence when it became caught up in this set of cultural evolutionary dynamics, and was co-opted to perform several novel functions linked to social norms and ethnic boundary markers. This will give us a clearer perspective from which to consider the questions raised by moral disgust. In section 4, we will address some of those questions, and formulate a byproduct hypothesis to account for some of the more puzzling features of the operation of disgust in domains linked to morality.

4.2 Developing the Tribal Instincts Hypothesis

Broadly speaking, gene-culture coevolutionary theory (GCC hereafter) seeks to understand the systematic interactions between innate, genetically specified information and the phenotypic characteristics it specifies, on the one hand, and the dynamics surrounding the social transmission of cultural information, on the other hand. Obviously, this is no small task. As one might expect, the number of issues on which the GCC literature touches is enormous (see Boyd & Richerson 2005). In order to pick a line through this work, we will be guided by our ultimate aim of illuminating the relationship between disgust and morality. After sketching the basic outlook and fundamental assumptions of the theory, we will clarify the more specific idea that one result of our immersion in culture has been that humans are now innately disposed to see their social world in tribal terms, and to react accordingly. In unpacking what, exactly, this idea amounts to, we will discuss the importance of imitation, social norms, ethnic boundary markers, and their impact on human cognitive architecture.

4.2.1 Gene-Culture Coevolutionary Theory

Gene-culture coevolutionary theory is sometimes called dual inheritance theory. It sees genetic information and cultural information as constituting two distinct inheritance systems, two structures that allow the transmission of information from one generation to the next. Genetic information is, of course, encoded in genes and transmitted biologically. Alternatively, cultural information is information stored primarily in brains[24], and passed from one generation to the next (as well as between members of the same generation) via many forms of social learning. Moreover, GCC holds that each type of inheritance system is subject to similar evolutionary forces, and that selective pressures shape the contents of each over time. Perhaps most significantly, it also sees the operation of the genetic inheritance system and the cultural inheritance system, respectively, as exerting systematic long-term influence on the operation and evolution of the other.

These interactions, according to the theory, have had a profound influence on human psychology. Beginning from many of the same premises and assumptions of classical evolutionary psychology, GCC supplements that theoretic outlook with the insight that humans are not just highly social but uniquely reliant on culture.[25] Sociality of any type requires some degree of communication and information transmission between conspecifics. One of the most fundamental insights of GCC is that our increased

---

[24] Cultural information can be stored in other mediums as well, most notably artifacts such as books, computer disks, and so forth.

[25] Briefly, classical evolutionary psychology holds that much of the cognitive architecture of the human mind can be likened to a Swiss Army knife: both are composed of a number of distinct, specialized parts, and those parts individually serve different kinds of functions. It sees the mind as a collection of semi-autonomous, domain-specific mental mechanisms, each of which evolved in response to a specific, recurring adaptive problem faced by hominids during their evolutionary past. Each mental mechanism is fairly specialized, both in that it is functionally specialized to solve a specific adaptive problem presented by the physical or social environment, and in that it is activated by a special set of cues relevant to that problem. See Barkow et al. 1992, Pinker 1997, and Tooby and Cosmides 2005 for more detail).

reliance on this type of socially acquired information, in contrast with information transmitted genetically, radically altered the selective pressures involved in human evolution. These novel selective pressures, in turn, had their most pronounced impact on human psychology and cognitive architecture. In their own words:

> "Our framework, however, emphasizes the additional possibility that adaptation to rapidly shifting evolutionary environments may have favored evolved psychological mechanisms that were specialized for various forms of learning, particularly complex forms of imitation (Richerson & Boyd 2000a; Tomasello 1999). We call the action of these mechanisms cultural learning. The idea is that, at a certain point in our cognitive evolution, the fidelity and frequency of cultural learning increased to the point that culturally transmitted ideas, technological know-how, ethical norms, and social strategies began to cumulate, adaptively, over generations. Once this cumulative threshold is passed, selection pressures for social learning or imitation, and the requisite cognitive abilities, take off. A species crossing this threshold becomes increasingly reliant on sophisticated social learning (Boyd & Richerson 1996). The fact that humans in all societies depend upon locally adaptive, complex behaviors and knowledge that no individual could learn individually (through direct experience) in a lifetime, motivates such a theory."
>
> (Henrich et al. 2006, page 842)

Let us begin with culture and cultural evolution itself. Once again, GCC sees culture in general as a repository of information passed from one generation to the next. Rather than in DNA sequences, however, culture is epigenetic, encoded and stored in brains. It influences the behavior of individuals, but is transmitted via social learning rather than genetic material. Once certain conditions are met and a critical mass of information is reached, the body of cultural information itself begins to cumulate, so that eventually it includes more than any one person could learn via trial and error and individual problem solving in the course of a single lifetime. The repository of information is gradually modified, refined, and added to by members of subsequent generations, so that it contains the accumulated wisdom of many generations. Additionally, as the size of the cultural inheritance system balloons, cultural items are

increasingly in competition to survive into the future generations. As some cultural items prove more useful and compelling than others, they are more likely to be passed along and thus represented in the inheritance system. In this way, the contents of the entire, snowballing body of information becomes subject to various forms of selection, some of which stem from what is useful, others from what is compelling. In very general outline, this is the recipe for the evolution of culture. GCC theorists have developed an array of sophisticated game theoretic models and computer simulations to more precisely study the properties of cultural evolution under a variety of empirically plausible conditions (see Boyd and Richerson 1985, 2005).

On the other hand, the presence of this cumulative body of culture and the reliance on socially transmitted information also generates a unique set of pressures on the human beings who rely on it. As the size of the cultural inheritance system balloons and its import increases, new pressures are created that select for psychological capacities allowing individuals to easily *access* and *use* information stored in that epigenetic pool of information. Once the body of cultural information is large enough and reliance on it becomes sufficiently high, the coevolutionary threshold or tipping point is crossed, and a feedback loop is generated. On one side of this feedback loop, the features of genetically specified psychological mechanisms allowing access and transmission exert influence on the evolution of the body of cultural information. On the other side, statistical regularities in the contents of the cultural inheritance system exert influence on the evolution of the psychological capacities required to make use of culture and culturally transmitted information. This *core coevolutionary feedback loop* at the heart of GCC provides a very general picture of the ways individual cognition and population level

processes can mutually influence each other over evolutionary time. It is represented in figure 4.1 below.

Cognitive Architecture and Dual Inheritance Theory

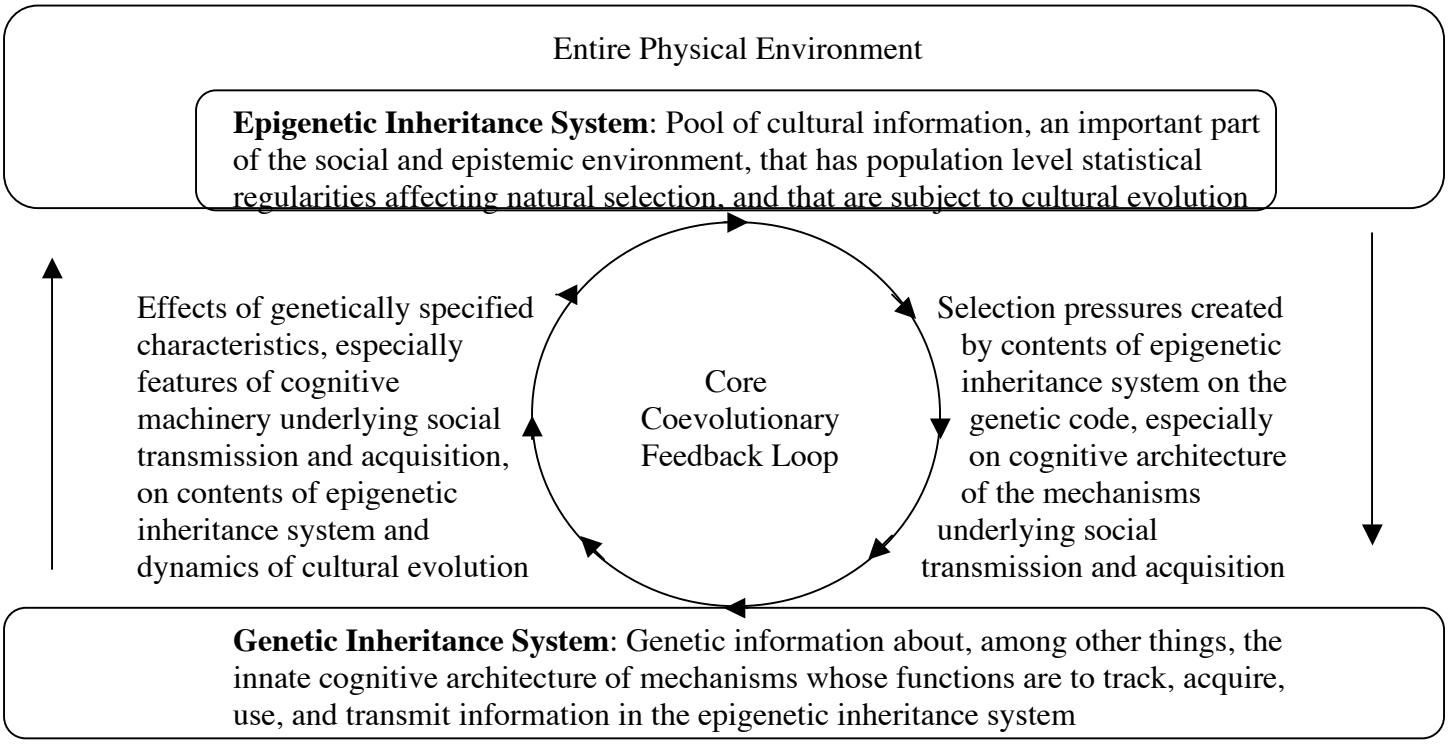


Figure 4.1
Innate, genetically specified cognitive mechanisms allow humans to access and manipulate information in the cultural repository. In virtue of mediating between the two inheritance systems, psychology is central to the study of cultural evolution, and the study of cultural evolution is just as important to psychology, especially in understanding the features human psychology that make us distinct and unique.

One key implication of GCC is that, due to its crucial role in mediating between individuals and the repository of cultural information on which they so heavily rely, human psychology is caught right in the middle of the major dynamics of gene culture coevolution.

How might this feedback loop change human psychological structure and the social dynamics it supports? Of course, a wide variety of changes might be expected – and there are probably even more that are not expected – but only a precious few of them

have begun to be explored in much detail.[26] Here we have only given the most general sketch of the GCC literature, which is large and complicated, but it should be enough to suggest how the novel perspective it affords on human evolution is bristling with insights and implications.

4.2.2 Tribal Instincts and Cognitive Architecture

Though the tribal instincts hypothesis was advanced by those whose main concern is population level dynamics and the evolution of culture, this component of the overall theory is best interpreted as a claim fundamentally concerned with human psychology, more specifically with features of human cognitive architecture. It maintains that one important consequence of the enfolding of cultural and natural selective pressures has been the evolution of a set of social "instincts" that are unique to humans. These instincts are sensitive to particular types of cultural information, namely types of information that structure and facilitate living within the context of large, cooperative groups or tribes. These instincts also lead to distinctive kinds of behaviors and inferences that are appropriate to tribal living. Given this collection of hypothesized features, tribal instincts are best interpreted as being subserved by dedicated cognitive mechanisms that process the information and cues to which they are sensitive in characteristic, perhaps biased and idiosyncratic ways.

Interpreted in this way, the perspective afforded by the tribal instincts hypothesis can be a source of genuinely novel predictive hypotheses about the types of features psychologists should be on the look out for when investigating the proximate

[26] Some of those attempts can be found in Sperber 1996, Boyer 2001, Atran 2002. For more discussion see Kelly et al. forthcoming.

mechanisms underlying social interaction.[27] But the tribal instincts hypothesis is also – and perhaps more fundamentally – concerned with *ultimate* explanations. A major source of motivation for this hypothesis is the claim that the complexities of human cooperative behavior, especially our propensity for living in large groups or tribes, outstrip what can be explained using only the resources available to the more widely endorsed varieties of ultimate explanations couched in terms of kinship or reciprocity (Richerson & Boyd 1999). Instead, GCC holds that what is sometimes called human *ultrasociality* is greatly facilitated by the fact that human social interactions are regulated by complex systems of norms, and that humans are able to recognize and selectively interact with members of their own tribe or ethnic group, who abide by the same set of norms. The tribal instincts hypothesis holds that a) there are dedicated cognitive mechanisms underlying different features of these abilities, and that b) any viable ultimate explanation of those cognitive mechanisms and the role they play in social cognition will crucially involve selective pressures generated by the core coevolutionary feedback loop. In their own words:

> "We believe that the human capacity to live in large-scale forms of tribal social organization evolved through a coevolutionary ratchet generated by the interaction of genes and culture. Rudimentary cooperative institutions favored genotypes that were better able to live in more cooperative groups."
>
> (Boyd & Richerson 2005, page 263)

Pressure generated by the core coevolutionary feedback loop favored individuals able to easily pick up culturally transmitted information – specifically information that facilitated living in large groups.

---

[27] For example, see McElreath et al. 2005 and Henrich et al. 2006 for preliminary attempts to experimentally test empirical predictions specifically derived from the GCC perspective.

Since work in gene-culture coevolutionary theory tends to focus on population level phenomena rather than the fine-grained features of cognitive mechanisms, GCC theorists rightly admit that nothing of precision can be derived about cognitive architecture directly from the tribal instincts hypothesis, without help from other disciplines like psychology, anthropology, and even archeology. As they put it, "the division of labor between innate and culturally acquired elements is poorly understood, and theory gives little guidance about the nature of the synergies and trade-offs that must regulate the evolution of our psychology" (Boyd & Richerson 2005, page 264). In the form of the tribal instincts hypothesis, however, the GCC perspective can provide a valuable and rigorous theoretic supplement to work done in experimental and comparative psychology and cultural anthropology. The coevolutionary perspective does have some broad implications for the type of cognitive machinery we should expect humans to come equipped with, given the selective histories that shaped them. By clearly articulating the adaptive problems that such mechanisms evolved to solve, we have a better sense of what such mechanisms would look like, how they might function, and what they might be doing. Below, we will tease apart three different aspects of human cognitive that the tribal instincts hypothesis suggests will be especially significant.

4.2.2.1 Imitation and Biases

It goes without saying that social or observational learning (roughly, learning by observing, retaining, and replicating the behavior of others) will loom large in the acquisition and transmission of cultural information. Given this central role, the instinct to imitate others is very likely to be well developed in species reliant on culture. Imitation or mimicry is one way of extracting information from the behavior of others,

and often leads to similar behavior. It could also lead to imitators entering more directly into mental states that are type similar to the mental states of the models they are imitating. Much recent work emphasizes the importance of psychological mechanisms associated with theory of mind (Tomasello et al. 1993), but mechanisms involved in empathy and emotion are likely to be involved as well (Hatfield et al. 1994).

Suggestive as this is, the relationship between imitation, learning, and cultural evolution, is still not well understood. Comparative research has begun exploring whether capacities for genuine imitation are necessary and/or sufficient not just for observational learning and the development of culture, but also for sustaining cultural evolution that is cumulative in the sense mentioned above (Heyes 1993, Boyd and Richerson 1996).

In addition research on imitation itself, preliminary empirical evidence supports mathematical models that suggest the human propensity to imitate is supplemented with a number of innate or "instinctual" biases. Members of one important class of biases are *context biases*. These predispose people to find compelling, and thus imitate, certain behaviors and attitudes found in a population based on who else has adopted them. These biases, by influencing which behaviors and attitudes will be most imitated, have substantial effects on which variants will be more widespread over an entire tribe or population (Boyd & Richerson 1985, 2005). Two such biases stand out: *prestige* biases lead imitators to embrace variants adopted by prestigious members of their culture (Henrich & Gil-White 2001). *Conformity* biases lead imitators to embrace those cultural variants most common among their peers (see McElreath et al. 2005 for experimental evidence on conformity). According to GCC, conformity biases in particular are

especially important in the emergence and maintenance of differences between groups (Henrich & Boyd 1998). Context biases are so-called because they are sensitive to features of the local social context rather than the intrinsic features of the behavior or cultural variant being imitated. While context biases can produce measurable effects on the population level distribution of cultural variants, they are relatively blind to the content of the variants they help propagate.[28]

Imitation can serve as a channel for the transmission of all sorts of cultural variants, including those directly connected to tribal living, but other sorts as well. Other instincts, however, might contribute to the transmission and cognition of cultural information pertinent to the more important features of tribal living. Indeed, a few more concrete suggestions about such instincts and the mechanisms giving rise to them can be extracted from GCC and the tribal instincts hypothesis. More specifically, two clusters of issues that are central to human ultrasociality and tribal living are *social norms* and *ethnic boundaries*.

4.2.2.2 Social Norms

Discussion of social norms spans many disciplines within the social sciences, and the term itself can have different nuances in the hands of different researchers. However, most would agree to a rough first approximation that characterizes social norms as rules regulating behavior and governing social interactions. Coevolutionary theory in particular sees norms as playing a crucial role in allowing humans to successfully cooperate in tribal sized groups. Norms also contribute to the unique ability of humans to

[28] In contrast, members another class of biases that can produce measurable effects on the distribution of cultural variants are sensitive to the content of the variants themselves; these are called content biases. Content biases come in a variety of forms. There does not appear to be any systematic relationship between content biases the tribal instincts hypothesis, however.

adapt to a wide range of environments (for more on the importance of social norms in coevolutionary theory, see Boyd & Richerson 1992, 2005; Richerson & Boyd, 1998, 1999).

Since psychological work relevant to social norms is scattered, we will begin by focusing on an account that was much inspired by the GCC perspective. In what can be read as a development of one strand of GCC's tribal instincts hypothesis, Sripada & Stich (forthcoming) focus on social norms, and argue that their escalating importance lead to the evolution of a set of innate, dedicated cognitive mechanisms underlying *norm psychology* in human beings.[29] In constructing a model of those proximate mechanisms, they emphasize many population level properties of norms, including the fact that norms are ubiquitous and important in all known cultures, often possessing a normativity independent of any external institution or authority. Sripada & Stich are also impressed by the fact that while the stability of all norm systems is supported by the punishment of violators, the norms that make up particular norm systems exhibit a pattern of within culture uniformity and cross-cultural diversity. Finally, while all members of a cultural group acquire the norms of the group in which they grown up despite any differences in their biological heritage, the specific behaviors prohibited by norms vary greatly from one cultural group to the next.

These group level properties of norms are accompanied by individual level properties, which are more directly relevant to the psychological focus of the tribal instincts hypothesis and Sripada & Stich's model. They maintain that the pattern of

---

[29] Other accounts of the psychology of norms and social rules have been offered (Nichols 2004, Prinz forthcoming). I focus on the S&S theory for expository purposes not only because it is admirably detailed, but because it was motivated, in large part, by the same types of findings and data that are crucial to gene culture coevolutionary theory.

within group similarity and cross-cultural diversity suggests that by and large norms are learned from the social environment. Their model thus posits a set of mechanisms dedicated to socially acquiring norms from others. Corresponding to the independent normativity and punishment supported stability found at the group level are the motivating effects that norms have on those individuals who have internalized them. Sripada & Stich assemble evidence suggesting that once an individual has acquired a norm, she thereby acquires the paired intrinsic motivations to both *comply* with that norm, and *punish* those who violate it. The model thus posits another set of execution mechanisms that produce those paired motivations.

Together with a database for storing the norms, the interlocking set of mechanisms dedicated to acquisition and execution make up the norm psychology. While those mechanisms are innate and universal, the evidence suggests that there can be considerable variation in the sorts of rules that can be acquired, in terms of the types of behaviors that they govern and the types of people to whom they apply. Indeed, different norms can apply to different groups of people; some may apply to everyone, while some may apply only to very narrowly circumscribed categories of people, such as children, adult women, unmarried men, members of one's own tribe, and so on. Further, norms can differ greatly with respect to the sort of punishment that will be directed at violators of the particular norms.

While the architecture posited by this psychological model is able to explain some systematic cross-cultural regularities in the cognition of social norms, it also allows for a high degree of diversity across tribes and cultures. More broadly speaking, the GCC perspective that inspired it can accommodate and begin to explain diversity in norms. In

the first place, different types of norms and norm systems are required to produce behavior that is locally adaptive to different environments. In addition, diversity can also be explained by appeal to the possibility of multiple stable equilibria: in some conditions there may be number of different possible norm clusters upon which the dynamics of a group might settle. Each equilibrium point or potential cluster is both internally stable and can produce relatively adaptive collective behavior in that single environment (Sripada 2005, see also Boyd & Richerson 1992, Henrich et al. 2006).

Sripada & Stich's proposal takes the form of a general framework, and many more specific questions remain open, including questions about the extent and nature of constraints on the types of norm that can be acquired, the proximal cues in the environment to which the acquisition and execution mechanisms are sensitive, and the representational format of norms. More detailed psychological research on norms or mentally represented rules, their representational format, and associated cognitive architecture is very much in its infancy (Nichols & Mallon, forthcoming). One question that has drawn much attention recently is the role of emotions in the psychology of norms and in moral judgments more generally (Nichols 2004, Haidt 2001, Greene & Haidt 2002, Prinz forthcoming,). However the details may turn out, what seems beyond question is that in some way or another, the mechanisms associated with the norm psychology often interact and work in conjunction with emotions and other psychological systems.

As for the ultimate origins of the psychological norm system, many suggestions have been made, but no systematic explanation has been advanced yet. The tribal instincts hypothesis promises to loom large in any viable candidate explanation, however.

Considering that the central structural features of the norm psychology are likely to be found universally amongst humans, and that they are also found only in humans, it appears that explaining their origins will require resources beyond those available to explanations couched solely in terms of kinship and reciprocity. If the tribal instincts hypothesis is correct, a viable explanation will appeal to the core coevolutionary feedback loop.

Indeed, the GCC perspective has inspired many of the most promising ideas concerning the evolution of our distinctive capacity to cognize norms. These include the centrality of punishment (Sripada forthcoming), cultural group selection (Boyd & Richerson 2005), the adaptive flexibility of the cultural inheritance system relative to the genetic inheritance system, the power of norms to fine-tune behaviors so that they are locally adaptive, given the contingencies of the different environments (Henrich & McElreath 2003), and the power of a system of mutually shared norms to stabilize and coordinate interactions in large groups (see Sripada forthcoming, Machery & Stich forthcoming). At this point, it is not clear how these proposals are related to each other – which are mutually incompatible, which might be complementary, etc. What is clear, however, is that each is intriguing in its own right, and deserves further investigation.

4.2.2.3 Ethnic Boundaries

A final strand of the tribal instincts hypothesis begins with the insight that large-scale sociality is further enhanced if actors to can make informed decisions about the individuals with whom they choose to interact. One way these decisions can be informed is if people can readily recognize those who adhere to similar social norms. In short, "symbolically marked groups arise and are maintained because dress, dialect, and other

markers allow people to identify in-group members" (Boyd & Richerson 2005, page 99). Here again, GCC provides insight into both the nature of the adaptive problem generated by the need to discriminate amongst potential interactants, as well as the nature of the strategies and cognitive mechanisms that solve, or at least mitigate, the difficulties caused by those problems.

The potential adaptive value of symbolic markings is perhaps less obvious than that of social norms. Discussions of ethnic boundary markers often begin with observation that humans in nearly all known regions and time periods have divided themselves into something resembling ethnicities. That is, humans have organized themselves into groups with which they identify, and whose members mark themselves with arbitrary symbols of various sorts (McElreath et al. 2003, Henrich & McElreath 2003; see also Barth 1969). Such a striking fact deserves an explanation. The deepest insight that GCC has to provide on this phenomenon is that ethnic marking is fundamentally linked to the fact that social norms govern interactions between group members.

In short, ethnic symbols allow members of the same "tribe" (or "ethnie" as they are sometimes called), to identify and selectively engage in interactions with each other. Why is this significant? Members of the same tribe, almost by definition, share a large set of beliefs, values, and most importantly, large clusters of social norms.[30] Sharing the same norms, in turn, facilitates *coordination* in those social interactions that are governed by them: actors will have similar and complementary expectations about the "proper" form of the interactions, practices, and customs in which they might mutually engage.

---

[30] See Gil-White (manuscript a) for a discussion of the vexing terminological difficulties here.

Coordinated interactions, in which the norms and expectations of the actors are matched, will go more smoothly, to the relative benefit of all parties involved. Alternatively, actors who don't share norms will often find that they are at odds with each other, or at least that their expectations are not aligned. As a result their behaviors will fail to mesh. This of course disrupts the interactions, to the relative detriment of all parties involved.

Thus, on this view, the function served by the arbitrary ethnic symbols is to maximize coordinated interactions. They do this by providing a visible, external, and physical signal of an underlying set of invisible, internal psychological dispositions, namely the beliefs, values, and clusters of social norms endemic to one particular tribe rather than another. The visibility of the symbols, of course, provides easily accessible information that helps all parties selectively engage in coordinated interactions, while avoiding those that promise to be uncoordinated and difficult. Since those symbols mark a set of psychological differences between members of one ethnic group and the next, they are called *ethnic boundary markers* (McElreath et al. 2003).

A noteworthy feature of this account of the function of tribal markings is that it does not directly appeal to altruism or cooperation (cf. Kurzban et al. 2001). By emphasizing coordination, rather than cooperation, it is able to sidestep many of the familiar problems associated with freeloaders and defectors, including those associated with costly and false signaling. On the one hand, this account has it that ethnic boundary markers *do* allow actors to be selective about whom they interact with. On the other hand, those markings do not purport to provide information about where to direct *altruistic* impulses, or which potential interactants are likely to *reciprocate* such impulses in cooperative ventures. It is easy to see how such a scenario would be unstable. If a set

of ethnic markings advertised indiscriminate cooperative tendencies, defectors and freeloaders could easily infiltrate a tribe of altruists by adopting a set the relevant set of markings, and reaping the benefits of others' altruistic behaviors without ever reciprocating, thus without ever paying the cost.

Rather, in facilitating *coordination*, ethnic boundary markers help maximize a feature of interactions that benefits *all* parties. Hence, unlike other signaling strategies directly related to cooperation and reciprocation, the information signaled by ethnic boundary markers provides no immediate opportunity for one actor to asymmetrically exploit another, without thereby diminishing her own returns. On this first approximation of the underlying social dynamics, then, there is little incentive to display false signals by adopting the ethnic markers of an unfamiliar tribe – it would be self-defeating.[31]

Once again, this appeal to social norms and resources besides reciprocity and altruism illustrates how the account of ethnic boundary markers is of a piece with the tribal instincts hypothesis. In describing a model of the evolution of ethnic boundary markers, Henrich and McElreath make explicit the link to the core coevolutionary feedback loop that creates tribal instincts:

> "The model makes predictions about both evolved psychological propensities and sociological patterns, and explicitly links them. Ethnic marking arises as a side effect of other psychological mechanisms—which themselves have solid individual-level selective advantages—that happen to generate behaviorally distinct groups. The strategy of using arbitrary symbolic markers to choose interactants then evolves because of features of the culturally evolved environment. Cultural transmission mechanisms may create statistically reliable regularities in the selective environments faced by genes. Thus, explaining many important aspects of human psychology and behavior will require examining how

[31] This is the case initially, at least. Once ethnic boundary markers have arisen, they are liable to become interwoven with social dynamics involving moral reciprocity, punishment, and cooperation. Moreover, certain markers could come to be associated with clusters of prosocial norms that recommend altruistic behavior. This, in turn, would provide an incentive for freeloaders to mimic them. See McElreath et al. 2003, page 128 for brief discussion and further references.

> genes under the influence of natural selection responded to the regularities produced by culture. This means that understanding the behavior of a highly cultural species like humans will sometimes demand a culture-gene coevolutionary approach."
>
> (Henrich & McElreath 2003, page 133)

These considerations suggest that in addition to a norm psychology, human tribal instincts will also include an evolved *ethnic psychology*.

To date, less research has been done on the specific cognitive mechanisms associated with ethnic psychology than on norm psychology. Here again, however, the GCC perspective affords valuable insight. For example, it implies that the importance of identifying and classifying ethnic actors as such generated selective pressures for dedicated mechanisms that were particularly sensitive to ethnic boundary markers. Gil-White (2001) suggests that here the ethnic psychological system borrows some of the same mechanisms that underlie folk biological categorization and the representation of species. Evidence suggests (Medin & Atran 1999) that these mechanisms initially arose to process information about biological entities, and as described by Henrich & McElreath above, the mechanisms themselves already had "solid individual-level selective advantages". According to Gil-White's proposal, they were then further co-opted to perform some of the functions associated with ethnic categorization. Gil-White argues that folk biological capacities provided a fit candidate to be co-opted to this purpose: they antecedently applied to living organisms, and produced inductive generalizations about unobservable properties of those organisms based on their observable properties. In the case of ethnic categorization, inferences about behavioral dispositions and social norms needed to be made on the basis of visible ethnic boundary markers.

One upshot of this proposal is that it allows for the explanation of some of the more idiosyncratic features of ethnic cognition. Since information about both species and ethnies is processed by the same mechanisms, unobservable attributes *in addition* to behavioral dispositions and social norms are projected onto actors based on their ethnic categorization. Strange as this may sound at first blush, there is evidence that suggests ethnic actors are "essentialized". Ethnic groups are cognized as if they shared many properties with biological species, and inferences made about ethnic actors suggest that the conditions for inclusion in an ethnic group include possession of an unseen, inner "essence" that is transferred biologically from parent to child. This alleged essence is cognized as if it outstripped any observable signs, and is also thought to underlie certain characteristic inferences associated with essentialized thinking (Gil-White 2001, see also Gelman 2003). Many of these inferences are clearly false in the case of ethnic actors, but they are easily explained by this proposal: they are byproducts carried over from the original function of the folk biological mechanisms into the new domain of ethnic identification and categorization.

Whether more research vindicates the details of this account of an ethnic recognition mechanism, it is unquestionably correct in its assertion that ethnic boundary markers and symbolic markings are salient to human actors. Another truism about symbolic markers and boundaries is that they are often highly motivating and emotionally charged. Gil-White's proposal leaves this aspect of ethnic psychology unaccounted for, however. More specifically, while it explains features of the identification and categorization of ethnic actors, the proposal remains silent on the types of motivation and emotional reaction characteristically produced by ingroup and outgroup members,

respectively. Some preliminary experimental work has been done investigating ingroup biases (Tajfel et al. 1971, Turner 1984, see also Richerson & Boyd 1998 for discussion) but to date no hypotheses have been advanced about what proximate psychological mechanisms produce those biases.

Of course, being sensitive to and inclined towards ingroup members entails being sensitive to but disinclined towards outgroup members. The darker side of ingroup preference and tribal solidarity is xenophobia, ethnocentrism, and prejudice. GCC locates these phenomena in an evolutionary context; in doing so, it puts a peculiar twist on them. Recall that from the point of view of coordination, interactions with outgroup members are likely to go less smoothly than interactions with ingroup members. Because of this, interactions between members of different tribes, who don't share social norms, will be costly to *all* parties. This, in turn, suggests that ethnocentrism, though clearly repugnant in many forms and largely at odds with moral codes founded on equality and egalitarianism, could very well be adaptive (see Bowles & Gintis 1998, 2001; see also Gil-White manuscript b for more discussion). Ethnocentric attitudes and instincts to avoid members of other tribes, triggered by their different or unusual ethnic markers, would decrease the number of uncoordinated and inefficient interactions – again, to the relative benefit of all.

From the point of view of the tribal instincts hypothesis, this fact suggests that another important component of the ethnic psychological system will be cognitive mechanisms that produce and support ethnocentrism and bias against outgroup members. Such attitudes are certainly pervasive; as Boyd & Richerson put it:

> "[G]roups of people who share distinctive moral norms, particularly norms that govern social interactions, quite likely become ethnically marked. This suggests that ethnocentric judgments easily arise because "we the people" behave properly, while those "others" behave improperly, doing disgusting, immoral things, and showing no remorse for it, either."
>
> (Boyd & Richerson 2004, page 101)

Although the widespread existence of ethnocentric attitudes cannot be seriously disputed, systematic research on the cognitive mechanisms that might produce, process, and sustain those attitudes still remains in its infancy. Initial findings the support common sense assertion that ethnocentrism and ingroup solidarity is emotionally charged. Interestingly, Cottrell & Neuberg (2005) give preliminary evidence that links different emotions to the prejudicial attitudes directed at different outgroups. Thus it seems that many of the mechanisms associated with the ethnic psychology will indeed have solid individual level selective advantages. Additionally, some of the most striking experimental work that has been done on biases strongly suggests that ethnocentric attitudes can take both implicit and explicit form. That research also shows that implicit ethnocentric attitudes are easy to acquire, difficult to eradicate or reverse once acquired, and that they require effort and attention to suppress (Greenwald et al. 1998, see Kelly et al. forthcoming for discussion).

4.3 The Social Character of Disgust

We now turn our attention back to disgust in particular. According to the Entanglement thesis advanced in chapter 2, neither of the mechanisms at the heart of the disgust execution system was primarily devoted monitoring social interactions, and neither arose from the Machiavellian push and pull of strategic interactions between conspecifics. Rather, one evolved to protect against poisons, the other against parasites. The behavior of other conspecifics would be somewhat relevant to the performance even of these capacities, however. Disgust got its foot in the door of the social world by way

of this antecedent sensitivity to others, and was thus able to acquire auxiliary functions that were more directly involved in regulating social affairs. Many of these involve the types of social interactions highlighted by the part of gene-culture coevolutionary theory that involves tribal instincts.

4.3.1 Primary Functions and Social Scaffolding

It will first be useful to consider how performance of the primary functions of the disgust execution subsystem would have been enhanced by the availability of social information, and how the need to better regulate food intake and disease avoidance would have begun selecting for mechanisms of social transmission and acquisition.

Recall that at the heart of the disgust execution subsystem are two distinct mechanisms, one that underlies acquired taste aversions, and whose function is to regulate food intake, and another that underlies a sense of offensiveness and contamination sensitivity, and whose function is disease avoidance. Together, these give disgust the primary roles of regulating food intake and protecting against potential poisons and parasites. Social interactions are of immediate relevance to the latter of these. Member of the same species are particularly salient to the problem of disease avoidance, in large part because microbes able to infect any particular member of a species are often transmitted via social contact with other members of that species. Thus, in performing one of its primary functions of avoiding pathogens and conspecifics who potentially harbor them, disgust is already in the business of regulating social interactions with other people, albeit in a very brute manner – namely by inhibiting them.

For instance, Kurzban & Leary (2001) explicitly locate parasite avoidance within the framework of social interaction and cooperation. They argue that once social

interaction and cooperation became crucial to human evolution and human (or human ancestor) life, so to did the need to be selective about whom one socialized with. This gave rise to the need to stigmatize some individuals:

> "The major point is that in order for sociality to be functional, there must be "brakes" on sociality. An organism that chose to socialize in any way with every other creature it encountered would be a strange one indeed and clearly at a selective disadvantage. We should expect therefore that natural selection would fashion constraints and limits on sociality that cause one to direct one's social efforts in productive ways. We suggest that these brakes, a result of the necessity to be discriminating in one's selection of partners for particular kinds of social interactions, might play an important role in generating the stigma phenomenon."
> (Kurzban & Leary 2001)

By inhibiting interactions with others who exhibit the marks of infection, disgust acts as one set of "social brakes".

It is not difficult to see how performing this function, in and of itself, could lead to pressures selecting for elicitor sharing and mechanisms for signaling between people. Others are antecedently salient to the disease avoidance mechanism. It is not only sensitive to parasites and pathogens themselves, but also to reliable indicators of their presence, such as the phenotypic abnormalities of others. Another indicator of the presence of parasites and pathogen, however, is the *behavior* of conspecifics, healthy or otherwise. If others systematically avoid a place, entity, etc., that very fact might indicate that it is contaminated. Indeed, while gleaning information about the environment from the behavior of others is a useful strategy in general, it is even more fitting in the case of detecting diseases. By using the behavior of others as a guide, a person need not get in close proximity, and thus expose oneself, to the potential source of infection. Instead, he or she can simply adopt the less risky strategy of taking another's word for it, metaphorically speaking.

There is even good reason to think that *individual level* selection would favor those who were more likely to "sound the alarm": to be disgusted, express their disgust, particularly with the relevant facial expression, and thus indicate to others the presence of disease. For, once living in a group, it is hard to see what would be gained by a person who was deceptive or secretive about any potential contaminants he had detected. Such secrecy, in the form of obscuring or withholding knowledge of a source of infection, would have the effect of allowing other members of the group become infected with a *contagious* disease. That contagious disease could then spread throughout the entire group, infecting and perhaps killing all of the members, including the original deceiver. Of course, dispositions to express disgust and sound the alarm would be worthless if they did not coevolve with corresponding dispositions to recognize the relevant behaviors and react accordingly. Here selective pressures shaping the expression and recognition mechanisms underlying disgust acquisition were also giving them a very rudimentary role in *information sharing*, and thus maintaining group cohesion, broadly speaking.[32] Though it is rudimentary, this is sort of role and disposition that Mother Nature is able to exploit and build on.

In protecting the gut against ingesting potentially toxic substances, disgust also plays an important role in regulating food intake. Like disease avoidance, the functions performed by the food intake mechanism can be enhanced by social information. Analogous pressures could have selected for an ability to signal and socially acquire information about toxic potential food sources. Shifting dietary habits contributed to a substantial overlap between the proper functions the two core mechanisms, and once

[32] For preliminary experimental evidence suggesting that disgust activates a bias towards information sharing, see Heath et al. 2001).

those two mechanisms of the execution subsystem had fused together, it appears they came to share the same signaling system. As noted above, these very pressures selecting for a signaling system probably played an important role in driving the two execution mechanisms together in the first place.

Additionally, the unique evolutionary pathway taken by humans also exacerbated the adaptive problems associated with disease avoidance and food intake in other ways. Our ancestor's turn down this "unique evolutionary pathway" appears to be very much bound up in our capacities for cultural transmission, social learning and tribal living.

4.3.2 Tribal Instincts and Disgust: New Adaptive Problems and Novel Functions

It will be helpful to briefly step back and get our bearings. At its most general level, the GCC framework describes how a body of social information can be sustained by groups of humans, and accessed and transmitted from one human to another. Moreover, it shows how reliance on such a body of social information created new adaptive problems and selective pressures, which in turn shaped the cognitive architecture of human psychology. These helped form a new set of "tribal instincts," according to the strand of coevolutionary theory we focused on. The novel adaptive environment, filled with radically new selective pressures generated by the core coevolutionary feedback loop, also created a set of conditions in which novel functions could be performed by ancient cognitive mechanisms, which had perhaps originally evolved for completely unrelated purposes. In other words, culture itself created an environment ripe for the *co-opting* of old cognitive mechanisms to new purposes.

The term *co-opt* is used here to capture the process wherein a preexisting trait or mechanism acquires a new function in response to novel or shifting selection pressures

from its environment.[33] The old trait might itself be an adaptation or not. If, however, it was an adaptation, it is possible for the new function to replace the old. On the other hand, it is also possible that the old function could continue to be performed alongside the new one, without impaired efficacy to either. In such a case, an auxiliary function is added to the primary function of the trait or mechanism, thus rendering it *multifunctional*. Unlike the process of descent with modification, co-opting involves little or no substantial alteration of the structure of the trait or mechanism. Rather, the emphasis is on changes in environment, and thus on the role trait or system is playing. The structure of the trait itself remains largely the same, while the novel selection pressures create a new niche, in which the trait acquires a new function. As is the case of descent with modification, most commonly discussed examples involve the co-opting of physical traits or characters. (A common example is insect wings, which initially evolved to preserve warmth, but gained the function of enabling flight once they were large enough.) Nothing in principle, however, prevents psychological attributes from being subject to the same process, or, for that matter, from being co-opted more than once.

With this in mind, let us return to the specifics of disgust, and develop the Co-opt thesis. We saw how the two core mechanisms of the disgust execution subsystem might take advantage of available social information to better perform their *primary* functions of avoiding diseases and regulating food intake. Moreover, the nature of those two mechanisms make the disgust system particularly susceptible to being co-opted,

---

[33] As opposed to the process, the *traits* that have themselves been co-opted are sometimes called preadaptations or exaptations. The term ”exaptation” is more often used when the trait in question was not previously adaptive, or functional at all (Gould and Lewontin 1979), while “preadaptation” is reserved for traits that performed some other adaptive function prior to being co-opted to play a new one (Mayr 1960). Though disgust is clearly an instance of the later of these, I wish to steer clear of the theoretic and philosophical baggage that has been built into these terms, and will avoid using either one.

especially to perform roles that involve regulating social interactions. Recall that on the one hand, disgust exhibits great diversity in its potential elicitor set, stemming from the flexibility of its acquisition system. In virtue of the salience of conspecifics to issues of disease and parasite avoidance, disgust was already in the business of monitoring social interactions. While the system was innately sensitive to "phenotypic abnormalities," those appear to be specified quite generally, as a flexible, open-ended set of initial guidelines that can be revised, refined, or augmented with information acquired from the environment.

In contrast to the flexibility of the acquisition subsystem, the disgust response itself is fairly consistent across elicitors and domains. One reason this is significant is that natural selection is sensitive to stable statistical regularities and correlations, and the response constitutes just one of these. Its rigidity makes the disgust response a reliably elicited, fixed action pattern, the type of prominent behavioral regularity that is visible to natural selection. Metaphorically speaking, the response became a standing option and type of motivation that was available when new functions arose that needed performing, or new adaptive problems arose that required solving. Disgust, then, consisted of a rigid, reliable type of motivation and behavior, paired with open-ended database of elicitors and a flexible acquisition system. As new adaptive problems arose, this combination of flexibility and rigidity made the disgust system ripe to be co-opted to new purposes, including purposes that little or nothing to do with food intake or disease avoidance.

When considered next to each other, a) the conditions created by the core coevolutionary feedback loop, and b) the nature of the disgust system, seem an almost ideal match for each other: the former generates a variety of new adaptive problems,

involving especially social interactions, and the later lends itself to being co-opted to deal with new adaptive problems, especially those involving social interactions. Moreover, independent selective pressures were driving the development of a nascent signaling that would have been extremely useful to a species becoming more reliant on socially transmitted information. Given this perfect storm of converging factors, it is not at all surprising that disgust has become as multifunctional as the Co-opt thesis maintains it has become in humans. While continuing to perform its primary functions it continued accruing auxiliary ones generated by the novel coevolutionary conditions. As such, disgust has become deeply entangled with our tribal instincts; indeed, appears to have become involved in various ways with all three aspects of tribal instincts discussed above: imitation, social norms, and ethnic boundaries.

4.3.2.1 Disgust and Imitation

Capacities for social learning are crucial to the ability to access and transmit cultural information, and the perspective of the tribal instincts hypothesis suggests that mechanisms producing genuine imitation are likely to underlie at least some of those capacities. That perspective makes a broad suggestion about the importance of imitation in general, and as mentioned, much of the fine-grained experimental work in cognitive psychology that has been connected to this aspect of cultural evolution has focused on theory of mind. The research on empathic recognition of disgust (and other emotion) expressions, however, provides another potential point of contact between theory and experiment. As we saw in the previous chapter, that research suggests that in cases of empathic recognition, the receiver is not only able to detect the disgust of others, but comes to enter an affective state that is type similar to the mental state of the sender;

recognizing disgust in another often involves becoming disgusted oneself. This is a particularly deep and direct type of mimicry. In coming to experience disgust herself, a receiver not only imitates the observable behavior of the sender, but thereby comes to mimic the internal, psychological state as well.

While the disgust signaling system may have been initially shaped by selective pressures specific to the primary functions of disease avoidance and food intake, those could have been supplemented by pressures associated with the core coevolutionary feedback loop, creating a mutually reinforcing set that further enhanced its potency. Additionally, more specific hypotheses can be teased out by placing the expression and recognition in the context of gene-culture coevolution. One set might explore the effect of context biases on elicitor acquisition. If this line of reasoning is on the right track, then those pressures generated by the core coevolutionary feedback loop would also have produced constraints and biases in the acquisition system. These would bias individuals to acquire some disgust elicitors rather than others. Individuals would be more likely to acquire disgust elicitors shared by the majority of the social group (conformist bias) and those of high ranking or successful members of the social group (prestige bias).

The fact that imitation, via expression and empathic recognition, appears to play role in the acquisition and transmission of social information about disgust does not immediately bear on its relationship to morality. There is reason to think that in providing such a powerful channel for transmitting information about disgust, imitation is also providing a channel along which information relevant to social norms and ethnocentric attitudes can be transmitted, especially those connected to disgust itself.

4.3.2.2 Disgust and Social Norms

In the case of social norms, and to a lesser extent ethnic boundaries, the tribal instincts hypothesis allows us to cleanly identify some of the specific roles disgust has been recruited to play in our moral psychology. From a psychological point of view, the disgust system can interact with the mechanisms that comprise the norm psychology in a number of ways. One somewhat general instance of this is provided by the Sripada & Stich proposal, which emphasizes motivations to comply with norms and motivations to punish those who violate them. As noted above, the precise role of emotion in moral judgment is still a matter of debate, but no theorists maintain that emotion plays no role at all. If Sripada & Stich proposal is correct about the paired motivations associated with social norms, this provides a pair of clearly specified roles that different emotions might play in at least some moral judgments. Different emotions can provide the motivation concerning compliance with different norms, either in the form an impetus to actually complying or an impetus to judge that the norm should be followed. Alternatively, different emotions can provide the motivation concerning punishment, either in the form of an impetus to actually punish or an impetus to judge that transgressors of the norm are wrong, and should be punished.[34]

Disgust is available to fill either of these roles. There are many possibilities on how this could work, but there are features we might expect of norms that recruit disgust, rather than some other emotion, to provide motivation. For instance, norms regulating behaviors that involve intrinsically disgusting entities, such as the proper disposal of corpses or bodily wastes, or activities that are antecedently salient to the disgust system,

[34] For obvious reasons, much work in experimental moral psychology does not involve actual transgression and punishment, but rather involves asking subjects for their judgments when given vignettes about transgressors and other types of moral dilemmas; see Nado, Kelly & Stich forthcoming, and Doris & Stich (2006) for an overview.

such as dining practices, are probably more likely to engage the disgust system for motivational purposes. This could be the case for both compliance and punishment motivations. For instance, disgust could provide the motivation to *comply* with a norm that says to never eat food with the left hand, which is reserved for body maintenance: the action itself would become aversive, and one would be motivated to avoid doing it. Disgust could also provide the motivation to *punish* those who violate the norm: the violator would be ostracized, avoided, considered dirty and contaminated, even gaped at.

Some very interesting work in psychology makes evident how breaking certain basic norms, or even merely *considering* violating those norms, or remembering an unethical act that one has committed in the past, can trigger a disgust-like reaction, marked by a felt need to engage in symbolic cleansing or purification afterwards (Tetlock et al. 2000, Zhong & Liljenquist 2006).

This interaction between disgust and the psychological system that deals with social norms provides one way in which the central elements of the response can be elaborated in culturally specific ways as well. The disgust response is rigid enough that its central elements, including, most prominently, a gape (even if in the form of a microexpression), sense of offensiveness and sense of contamination, will be exhibited to some degree whenever disgust is triggered, regardless of other circumstances like the nature of the elicitor, other psychological systems that are activated, or even the culture of the actor experiencing disgust. However, culturally specific norms that utilize disgust might include more detailed information about the locally correct way to express the various elements of that response. In other words, while the clustered components are always produced in some form, social norms may help refine their expression. These

more finely honed displays can easily differ in specifics from culture to culture, but they will broadly instantiate a pattern of variations on the universal themes provided by the core disgust response (for instance, see Nemeroff & Rozin (2000) for discussion of local variation on the universal themes called the "laws of sympathetic magic").

Culturally specific norms governing disgust displays are often about social signaling more than anything else, but cultural information can help fine-tune norms that recruit disgust in other important ways as well. For instance, food taboos, broadly construed, provide one particular case in which the universal features of disgust work in conjunction with culturally specific information encoded in social norms that elaborate and enrich those universal features. From an evolutionary point of view, especially one supplemented with the resources of GCC, norms governing the practices surrounding diet stand out as class of social norms of singular importance. For a nomadic and omnivorous species such as humans, the problem of locating, obtaining, and preparing nutritional resources could take many different forms, as each environment provides a very different sets of dietary possibilities. A set of culturally transmitted, tribally specific, locally adaptive food taboos would be of prime value in navigating those possibilities. Moreover, such norms could help coordinate the collective efforts directed at location, procurement and preparation. These could refine and augment the rough guidelines provided innately by the disgust system. The need for this type of behavioral fine-tuning in the case of food consumption, as well as other locally adaptive practices that are directly linked to diet and nutrition – hunting strategies, foraging strategies, food preparation strategies, etc. – was paramount. Indeed, it could very well have provided

one of the most fundamental and significant pressures shaping the evolution of the norm psychological system itself.

Finally, the disgust system can also influence the population level distribution of social norms by providing content biases on their social transmission. With content biases, as opposed to context biases, the *content* of some cultural variants, rather than the social context in which they are transmitted, makes them more or less likely to be adopted and socially transmitted than others. The increased frequency of some variants is often explained by appeal to the properties and widely shared elicitors of individual psychological mechanisms. Disgust provides a specific instance of a content bias: agents are more likely to adopt and pass along cultural variants associated with disgust universals like phenotypic abnormalities, body fluids, decay, etc., probably because they are made salient to agents by their activation of the disgust system.[35] This content bias has been hypothesized to affect the evolution not just of single norms, but entire clusters of them as well, by influencing which of several locally stable equilibrium states a developing norm cluster settles upon (Shweder et al. 1997, Rozin et al. 1999).

4.3.2.3 Disgust and Ethnic Boundaries

Surprisingly, the most interesting overlap between disgust and morality has received the least systematic attention. Many have noted that disgust plays some role in marking and sustaining boundaries between groups, but other than bemoaning this fact, little is offered by way of clarification or explanation. Here, again, GCC and the tribal instincts hypothesis provide valuable insight and theoretic context.

---

[35] See Nichols (2002), Fessler & Navarrete (2003), and Heath et al. (2001) for examples of a disgust content bias.

To begin, the idea that nutrition, food taboos and norms regulating eating practices are of unique importance in the coevolution of our tribal instincts applies to the ethnic psychology as well. Different tribes, situated in different environments, will settle on different diets and clusters of food taboos. Behaviors related to cuisine – what food one will eat, what one is disgusted by and refuses to eat, how one procures and prepares that food, what methods of procurement and preparation one is disgusted by – provide a clear, observable source of information about the types of food taboos (again, broadly construed) that one adheres to. This information is about something quite basic to survival, but on the plausible assumption that clusters of dietary practices and food norms correlate with clusters of norms governing social interactions, eating practices provide information about the other types of social norms one accepts. In short, the many facets of cuisine come to act as obvious ethnic boundary markers.

Disgust, of course, is intrinsically linked to cuisine in virtue of one of its primary function regulating food intake. Given the tight connection between cuisine and ethnic boundary markers, it would have been a small step for the disgust system itself to be co-opted to play an important role in marking ethnic boundaries as well. More specifically, visible aspects of the disgust response like the gape face can themselves play the role of ethnic boundary markers, especially when elicited by particular types of food. When they reveal what disgusts them and what doesn't, ethnic actors show their colors. Being disgusted by an exotic food or practice, or alternatively, not being disgusted by some particularly odd food or practice, can itself mark whether one is a member of one particular group or another.

This idea, which already meshes with common sense and everyday anecdotal report, is made much more plausible when placed in the context of the tribal instincts hypothesis. Indeed, GCC provides resources to explain related cases of behavior that seem blatantly irrational on their face. For instance, when nutritional resources are scarce, being disgusted by, and thus refusing to eat an available type of food is clearly maladaptive. Such cases can be made sense of, however: the refusal to eat an available food source acts as an expression of commitment to a set of food taboos that forbid it. Being disgusted by some food, then, can be seen as a *costly* signal of one's membership to their tribe and its norms. A few striking instances of this have been discussed in more detail (Henrich 2001). Psychological experiments exploring the implications of this idea are still scarce, but there has been some preliminary work done (see Rozin & Segal 2003).

Once embroiled in the dynamics of ethnic boundaries marking, disgust appears to have been further co-opted. Disgust is the emotion, or one of the prominent emotions (fear being another likely candidate) of xenophobia, prejudice and ethnocentrism. As ethnic boundaries and ethnocentrism gained in adaptive value, the tribal instincts hypothesis predicts selective pressures would have driven the creation, or, more likely, recruitment of cognitive mechanisms dedicated to monitor and react to ethnic boundaries. One component of the ethnic psychology that has been hypothesized initially evolved to process information about species, but was further co-opted to be sensitive to ethnic boundary markers, and recognize ethnic actors as such. It seems evident that another such component is the disgust system. Disgust was available, and came to work in

conjunction with the ethnie recognition system, to provide the *motivation* to refrain from interacting with members of other tribes, once they are recognized as such.

In so doing, however, the operation of disgust appears to provide a propensity to demonize and dehumanize members of those other tribes as well. As in the case of social norms, recruitment of the disgust system entails recruitment of the entire cluster of elements making up the behavioral response. Some of those elements made disgust an excellent candidate for the purposes at hand – when members of other tribes trigger disgust, an actor is thereby strongly motivated to *avoid* them. But as noted above, disgust elicitors evoke the entire cluster of components comprising the disgust response, including the propensity to treat those elicitors *as if* there were offensive and contaminating, even when they are not. In instances where disgust underlies prejudices and ethnocentric attitudes, then, actors will not only avoid members of those tribes, but will be more likely to project offensiveness and contamination potency onto them as well, and will thus judgment them to be offensive, contaminating, unpleasant – in a word, disgusting. Those with different values and norms, members of other tribes that do things differently and give priority to different moral principles are not just different, but tainted, contaminating, immoral and somehow less or lower than one's own tribe – animal or sub-human.

Here again, the idea that disgust is responsible for such attitudes is completely in line with common sense and anecdotal reports, but psychological data supporting it is just beginning to be gathered. Nevertheless, what has been discovered thus far is compatible with the position adopted here. For instance, one study suggests that heightened disgust sensitivity correlates with xenophobic attitudes (Faulkner et al. 2004). Other work finds

that activities, and their perpetrators, that involve breaking some core norm or disregarding a central and defining value of the cultural ingroup are often considered not just wrong, but disgusting by members of that ingroup (Haidt et al. 1997). In terms of ethnic boundary markers, flouting a core social norm or defining value of a tribe, as opposed to some trivial norm or value common to many tribes, is another way ethnic actors might show their colors. Ingroup members may thus see such substantive transgressions not merely as isolated transgressions of particular norms, but also as violations of the entire tribe and the set of values that bind it together. Likewise, such violators can be seen not just as mere transgressors, but also as threats and outsiders of the worst kind, and thus appropriately shunned and worthy of disgust.

In addition, defenders of this the idea that disgust plays this type of role in moral judgments often point to well-known instances where one group is subjugated and dehumanized by another, such as the subjugation of the Jews by the Nazis, the attitudes taken towards members of the lowest castes in the traditional Indian caste system, or even less extreme instances of the dim attitude taken by an upper class towards a lower class. In such cases, the subjugating group often uses the idiom of disgust in characterizing the lower group as uncivilized, barbarian, animalistic and dirty (Miller 1997). Disgust can also have a particularly pernicious effect on such attitudes in virtue of the powerful but subliminal (perhaps powerful *because* subliminal) influence it can have on evaluations and more measured reasoning (Wheatley and Haidt 2005, Murphy et al. 2000, see also Haidt 2001). Neuroimaging research has recently begun to fill in some of the details. Not only has it confirmed the link between the most intense forms of prejudice and ethnocentrism and the brain areas associated with disgust, but it also confirmed the

correlation between disgust and *dehumanization*: only in cases of prejudice where disgust was the accompanying emotion did the higher brain areas associated with agency and interaction with other people (medial prefrontal cortex or MPFC) fail to activate (Harris and Fiske 2005) – when an outgroup member is disgusting, he or she isn't even cognized as a *person*!

4.4 Disgust and Morality: A Byproduct Hypothesis

Our concern in this chapter has not been with the normative issue of how or whether feelings of disgust should figure into our moral judgments. Rather, we have been pursuing the descriptive and explanatory goal of identifying the role or roles that disgust does, in fact, play in our moral psychology. In illuminating the connection between disgust and morality, and clarifying the roles associated with social norms and ethnocentric and prejudicial attitudes it has come to play, we have sought to fulfilled our final desideratum, and complete the theory of disgust. The 1st figure below reproduces the cognitive model given in Chapter 1; the 2nd depicts the same model, but with some of the more specific acquisition mechanisms and the evolutionary functions of the execution mechanisms filled in.

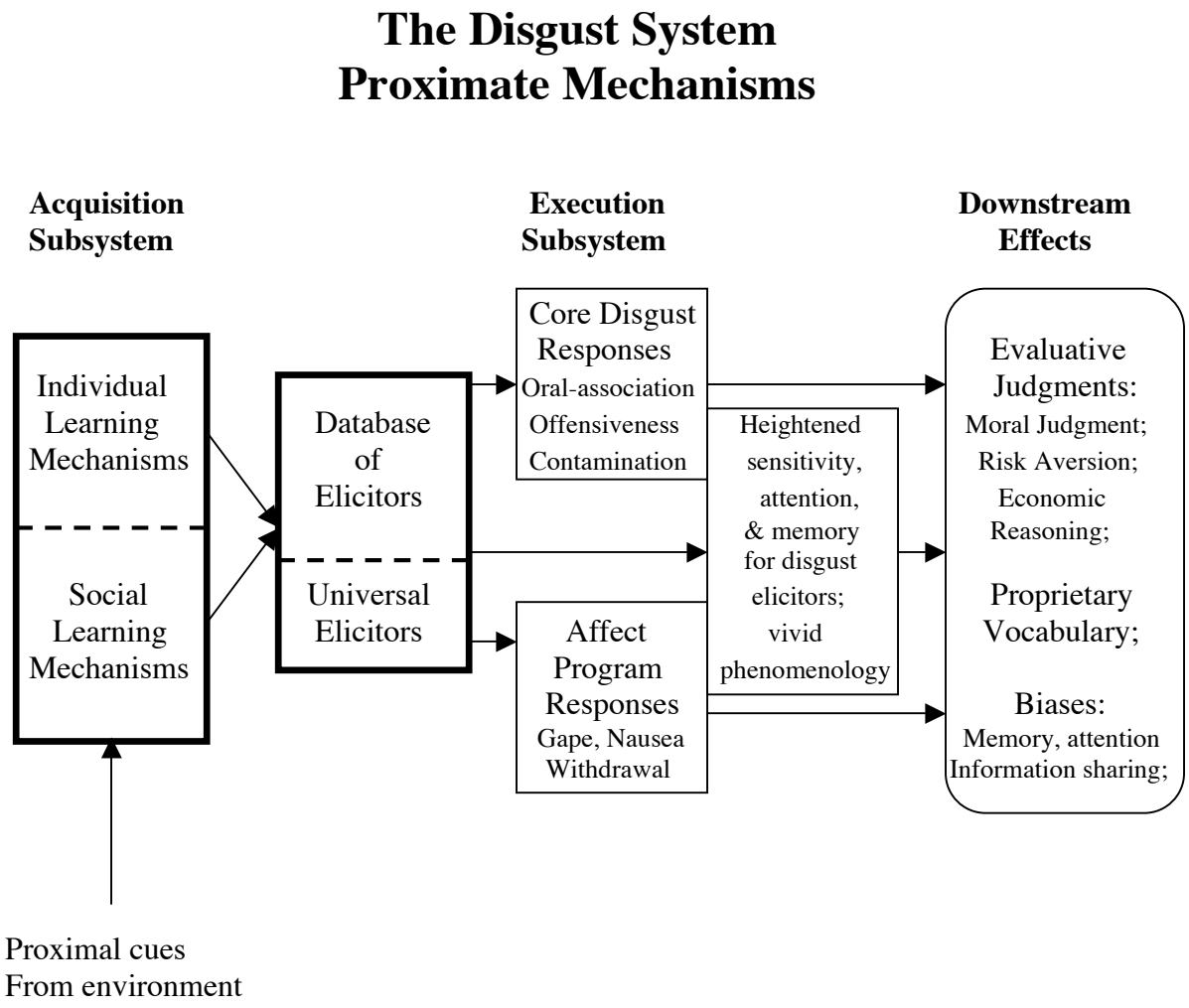
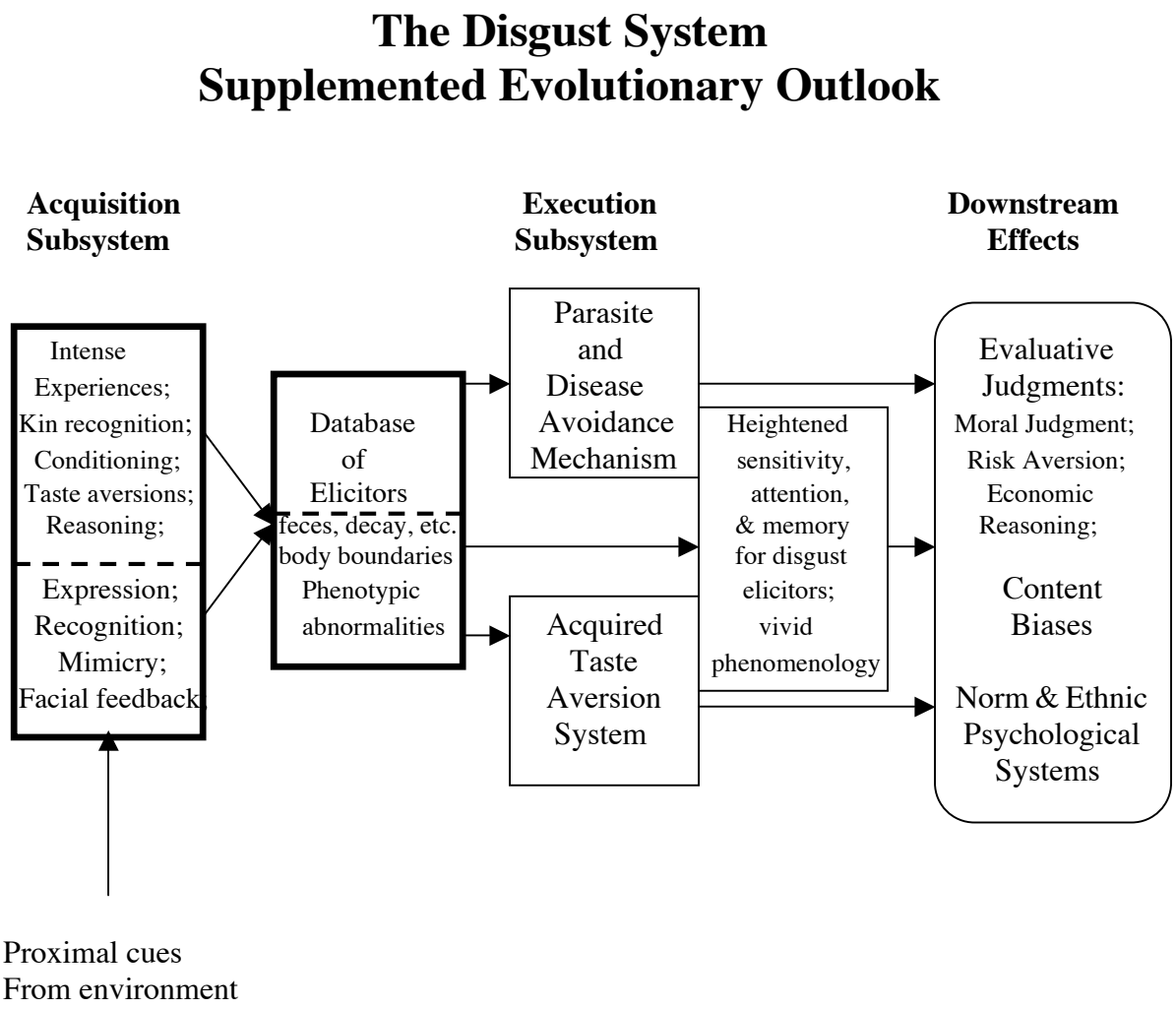


Figure 4.2
Functional level model of interlocking mechanisms that comprise the human disgust system. The arrows represent causal links between the various mechanisms.



Figures 4.3
Functional level model of interlocking mechanisms that comprise the human disgust system. The arrows represent causal links between the various mechanisms.

4.4.1 Demarcating the Domain of Morality

One of the questions that we asked at the beginning of this chapter was whether there is any substantial difference between instances of basic disgust and instances of moral disgust, and if so how that difference should be characterized. To the extent that calling something "moral disgust" requires that the domain or definition of morality be clearly delineated, this question remains difficult to answer. For, it remains unclear what determines, or how we should adjudicate, what properly falls within the domain of the

moral, especially when our project is descriptive. (However, see Nado et al. forthcoming for a brief discussion of various ways of construing the project of defining morality).

On the one hand, one might attempt to distinguish the moral domain from other domains using the methods of experimental psychology, demarcating norms that govern moral matters from religious edicts, conventional rules, or issues beyond the reach of morality that are located within a personal domain of autonomy. Developmental psychologists working on the so-called moral/conventional distinction can be seen as attempting to do this. They claim to have found systematic, cross-culturally stable differences in the way moral norms and conventional norms, and transgressions of those respective types of norms, are cognized along a variety of dimensions. On this view, moral norms are those having to do with harm, justice, welfare, or rights, and are judged to hold generally, rather than being situationally and culturally specific, and to hold independently of any authoritative figure or institution. Furthermore, transgressions of moral norms are judged to be more serious than transgressions of conventional norms (Nucci 2001, Turiel 1983).

There is reason to think that the impressive amounts of data gathered in support of this view paint a misleading picture, however. Rather, as my colleagues and I have argued at length elsewhere, the appearance of a sharply demarcated, cross-culturally robust divide between moral and conventional norms, so characterized, is an artifact, an illusion produced by the very circumscribed set of norms and transgressions used in the relevant experiments (Kelly et al. 2007, Kelly & Stich forthcoming). Whether other attempts to use the methods of psychology in determining which norms are, properly

speaking, the moral norms, remains to be seen, but the issue is vexing, and progress will likely be very difficult (though see Sripada, manuscript).

On the other hand, which individual social norms, judgments, and perhaps even which particular psychological systems should ultimately be classified as "moral" may not be a matter for science to discover. Rather, demarcating the moral might be a more conventional, cultural, and perhaps normative matter on which different cultures can each decide, and each justifiably decide differently (for further discussion, see Machery & Stich, manuscript, Shweder 1997). Even if it ultimately proves tractable, as long as it remains unresolved this difficulty will manifest just as it does in the more specific case of demarcating episodes of moral disgust from the rest.

Indeed, the roles that disgust plays in regulating social interactions, especially those associated with ethnocentrism and prejudicial attitudes, may not easily fit in some conceptions of morality at all, at least those that emphasize equality and egalitarianism. Such conceptions, in turn, seem to be most firmly entrenched in the view of morality held most dear by Western liberals. The tension and discomfort this is liable to create is nicely is captured nicely by one prominent social theorist:

> "The possibility of community is also weakened, in the long term, by the democratic principle of equality. If the strongest communities are bound together by certain moral laws that define wrong and right for its members, these same moral laws also define that community's inside and an outside as well. And if those moral laws are to have any meaning, those excluded from the community by virtue of their unwillingness to accept them must have a different worth or moral status from the community's members. But democratic societies constantly tend to move from simple tolerance of all alternative ways of life, to an assertion of their essential equality. They resist moralisms that impugn the worth or validity of certain alternatives, and therefore oppose the kind of exclusivity engendered by strong and cohesive communities."
>
> Fukayama 1992, page 323

This can be quite disorienting, and perhaps one more source of the sense of paradox some theorists have felt about this emotion. By the lights of the locally dominant, explicitly avowed moral code familiar to those of us in the West, one of the most prominent functions that disgust has come to play in our moral psychology is to support decidedly *immoral* attitudes!

We may be better served, then, by approaching the issues of morality and disgust by asking a slightly different, and equally interesting question: what specific role or roles does disgust play in the psychology that underlies judgments about putatively moral-oriented affairs, those having to do with how to conduct oneself, how to interact with others, and so forth? To answer this question, a distinction between strong and weak moral disgust is useful. As we have seen, all elicitors appear to activate the entire cluster of components in the disgust response, regardless of the nature of the elicitors. In some cases, disgust has an indirect, negative influence on judgments about moral issues, even when the emotion was triggered by something completely irrelevant to the moral judgment, such as the unsanitary conditions of the environment in which the judgment is made. From a psychological point of view, there is nothing specific to morality about this sort of disgust; downstream effects of the emotion have similar influences on other types of judgment and reasoning. Call this weak moral disgust.

However, we have also seen how the disgust system appears to work on its own when performing its primary functions of monitoring food intake and avoiding disease and parasites, but works in conjunction with other psychological mechanisms when performing many of its auxiliary functions – with a mechanism for kin recognition in the case of incest avoidance, with components of a norm psychology, with components of an

ethnic psychology. In these later cases, where the disgust system is paired with other psychological mechanisms, differences between "moral" and "basic" disgust might be traceable to the operation of the other mechanism, or features emerging from the interaction between the two. Call this strong moral disgust.

Though it has not yet been tested, there could be genuine, systematic differences between strong moral disgust and basic disgust. One way to operationalize such a distinction would be to extend a suggestion made in another context by Shaun Nichols (2004, chapter 1). Starting with the venerable distinction between merely being bad versus being wrong, we can go on to draw an analogous distinction between being merely disgusting (changing your child's nasty diapers, biting into a piece of spoiled meat that you left in fridge for too long) and being wrong (eating with your left hand in India, engaging in necrophilia). If some behavior is judged as merely disgusting, that suggests it is an instance of basic disgust, i.e. the disgust system acting on its own. If, however, some action is judged as both disgusting and wrong, that suggests that it is an instance of "moral" disgust, or disgust working in conjunction with the norm system. One way to get at this distinction more indirectly might be to ask, with respect to the particular behavior in question, whether anyone deserves to be *punished* because of it.

4.4.2 Cognitive Byproducts

We can end this section by juxtaposing the Co-opt thesis with the Entanglement thesis defended in chapter 2. Doing so allows us to explicitly formulate an idea that has been a tacit leitmotif of this chapter. As we have seen, once it is swept up in the social dynamics and selection pressures generated by the core feedback loop, disgust acquired several novel functions associated with regulating social interactions. These include

generating the motivations and proprietary attitudes associated with certain classes of social norms and certain types of ethnic cognition. At first, this might present a prima facie problem for the Entanglement thesis, which maintains that at the core of the disgust system are two mechanisms, one linked to poisons, the other to parasites. For, neither of those mechanisms have anything to do with morality, social norms, or ethnic cognition.

Rather than present a problem, however, we can use the Entanglement thesis to help further illuminate the way in which disgust informs these matters, and shapes the putatively moral judgments and motivations it does affect. Recall that according to that hypothesis, when the food rejection mechanism and parasite avoidance mechanism fused, they also created a system whose character made it highly susceptible to being co-opted to perform other functions, including functions pertaining to issues in the domain of "moral disgust" associated with social coordination and interpersonal judgment. As a result, it was able to acquire those auxiliary functions of providing motivation to comply with and punish violators of certain social norms and of influencing certain types of ethnic cognition and sustaining extreme prejudicial attitudes. The disgust *response*, however, remains highly consistent across all these new domains. Moreover, it seems much better fitted to its primary functions than to the auxiliary functions that it later acquired. For, in performing these novel functions the full nomological cluster of elements that comprise the disgust response was brought to bear on those social norms and prejudices. Hence, the behavior and attitudes driven by disgust in these new domains can be effective, but highly idiosyncratic, even irrational.

Evolutionary theorists often call explanations of this form *byproduct hypotheses*. Byproduct hypotheses are often advanced to explain a puzzling but systematic (rather

than random) deviation from optimal performance of an activity or function. The less than optimal performance is explained by appeal to the influence of some trait or system that is not performing its original function, but some new one. The systematic deviations are then explained as byproducts of the imperfect fit between the performance of the trait or system and the new function it has been co-opted to perform, or activity is involved with. Finally, the exact character of the systematic deviation is explained by appeal to specific features that the co-opted trait or system retained from its original function, and brought to bear on its new one.[36]

In the case of disgust, social norms that co-opt the emotion recruit a type of aversion, perhaps motivating agents to avoid the types of activities proscribed by those norms, or motivating them to avoid or shun transgressors. As a byproduct, however, such norms and motivations will also be infused with the other elements that accompany the disgust response, including a sense of offensiveness, contamination, and feelings of nausea. Thus the byproduct hypothesis can provide a preliminary explanation for the some of the idiosyncratic and irrational aspects associated with such norms, most notably the link to defilement and sanctity we find with many purity norms, the inclination to cleanse oneself after violating purity and other norms, and the extremities of the attitudes directed towards other transgressors. Likewise with the role that disgust plays in ethnic cognition: sensitivities to group membership might co-opt disgust to provide a powerful type of motivation, causing agents to avoid interactions with members of other tribes. While avoidance in itself might be adaptive in this context, motivation to avoid that is

---

[36] Other psychological byproduct hypotheses have been offered, for instance, to explain features of the character and persistence of religious beliefs (Boyer 2001, Atran 2002), aspects of ethnic and racial coginition (Gil-White 2001), and patterns of homicide involving male sexual jealousy (Daly & Wilson 1988).

supported by disgust will also inherit, as a byproduct, all of the key features of the disgust response. Once again, the byproduct hypothesis can provide a preliminary explanation of the gratuitous and irrational rhetoric about the contaminating, tainted, and less than human, "animality" of members of other ethnic groups that accompanies extreme cases of prejudicial attitudes.

More speculatively, but also more troubling, is the fact that feelings of disgust can induce judgments that are gratuitous and irrational, but nevertheless remarkably persistent. Recall the particularly vivid example discussed in chapter 1, that used a vignette about Dan the "popularity seeking snob": "Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He often picks topics that appeal to both professors and students in order to stimulate discussion (Haidt and Wheatley 2005). Those hypnotized to feel disgust at the word "often" judged Dan to be doing something morally wrong, and continued to endorse their initial negative moral judgment even when they were unable to provide credible reasons or justification for it. If the byproduct hypothesis is on the right track, similar persistence might be brought to bear on many other of the attitudes or norms involved with disgust (for instance see Haidt 2001).

4.5 Conclusion: A Uniquely Human, Multifunctional Cognitive System

The burden of this chapter was to begin illuminating the relationship between disgust and morality from a descriptive and explanatory perspective. The Co-opt thesis we developed shows, in broad outline, how disgust might have become as multifunctional as it has become. In doing so, it allows us to discharged that burden. By embedding the emotion within a the larger framework of gene-culture coevolutionary theory, and

developing the tribal instincts hypothesis, we were able to provide a theoretic context with which to make sense of much of the experimental data gathered about moral disgust, and to begin integrating the insights won from different approaches to the studying the emotion. Moreover, the perspective afforded by the Co-opt thesis, together with the Entanglement thesis about the origins of the disgust response, also allowed for the formulation of a byproduct hypothesis that can explain many of the more puzzling features of moral disgust.

Introduction to Part II: Shifting Gears from the Philosophy of Psychology to Metaphysics

In the next couple of chapters we will shift gears. Our primary concern will be with the types of questions that arise in metaphysics, broadly construed, specifically in what we called in the Introduction the *Metaphysical Project*. While our focus will change with the questions we are asking, however, we will not completely set aside the work we have done in the Empirical and Integrative Project. Rather, the completed theory of disgust constructed in the first part of the dissertation will afford us a new fresh perspective on the issues to be addressed in the second part. But first, let us step back and consider some of those issues, and how work from the one project might help illuminate work from the other.

Our experience of the world is often far from neutral. We may experience a sunset as beautiful, a rotting corpse as viscerally repellent, a colleague as charismatic and attractive, an act of heroism as moral, or a child molester as deeply disgusting. Intuitively, we accept that the world is as we experience it to be; we experience certain acts and entities as, for instance, valuable or repellent because they *are* valuable or repellent.

Philosophy begins in wonder. One of the issues it wonders about most obsessively is whether this intuitive acceptance of our experience at face value is a naïve mistake. Is the world actually as we experience it to be? Are beauty, value, and disgustingness real, part of the fabric of the universe, or is value merely illusory, beauty only in the eye of the beholder? How much of what we experience is properly ascribed

to the acts and entities in the world, and how much is added, projected onto those acts and entities by the perceptual and psychological apparatus that generates that experience? Questions such as these have a venerable philosophic pedigree, stretching from Plato's allegory of the cave, through Descartes' evil genius, Locke's distinction between primary and secondary properties, and Hume's observation that the mind has a propensity to spread itself onto the objects it observes, gilding and staining them with internal sentiment. Refined descendants of these concerns and ideas are still in high currency today, especially in contemporary metaethical debates between moral realists, quasi-realists, sentimentalists and error theorists.

Despite this long tradition, much of the philosophic discussion utilizes only the most rudimentary conception of the nature of the psychological apparatus likely to be involved. This, in no small part, was due to the fact that so little was known about that psychological apparatus; indeed, until relatively recently, psychology remained a highly speculative enterprise in general.

The last century, however, has seen a rapid maturation of the study of the mind. Advances in experimental methodology and technology have greatly enriched our ability to gather useful data, while advances in theory and formal techniques have greatly enriched our ability to identify patterns and make sense of that data. The cognitive sciences represent a large, loosely overlapping set of such techniques for investigating different aspects of the mind. Disciplinary boundaries separating those approaches remain, but the gradual convergence on a similar set of questions and issues has made the boundaries between the disciplines more porous, and amenable to integration. The

unification of the cognitive sciences – institutional but also, more importantly, conceptual – is still very much a work in progress.

Indeed, I take the work done in the first half of the dissertation to be a contribution towards this end. For, in the integrated, empirically rooted theory of disgust, we have an example of how conceptual resources and data gathered with a variety of approaches can be made to work together, helping explain different aspects of a single emotion.

This leaves us with questions arising in the Metaphysical Project. For, even in the unfinished form it is in today, the flowering of the cognitive sciences can provide powerful resources for addressing philosophic questions as well. Harkening back to the Introduction, we identified three different projects that researchers making claims about disgust have been involved in, the Normative, the Metaphysical, and the Empirical and Integrative. For the sake of clarity, it was and remains useful to separate each of these three projects, and keep the primary goals of each distinct from the others. However, it is also likely that in practice each project can, will, and should mutually inform the others. In general, the exact nature of the ideal interrelations between the three projects are not immediately clear, and in the case of specifically normative and metaethical questions, well-known concerns associated with the open question argument or the naturalistic fallacy are never far off.

Luckily, prominent authors have advanced a few broad guidelines that are difficult to deny. For instance, Owen Flanagan (1991) has offered his Principle of Minimal Psychological Realism: “Make sure when constructing a moral theory or projecting a moral ideal that the character, decision processing, and behavior prescribed

are possible for creatures like us." This is more or less an elaboration of the Kantian principle that "ought implies can", that a moral theory shouldn't ask us to do something that we aren't able to do, in this case something that is psychologically impossible for creatures like us. The issue of what is, in fact, psychologically possible or impossible for human beings is one that work in the Empirical and Integrative project can help shed light on. Additionally, Stephen Darwall, Allan Gibbard, and Peter Railton (1992) end their excellent survey of ethics at the close of the 20th century "Toward Fin de Siecle Ethics: Some Trends" by sounding a clarion call for more interaction between the Empirical and Integrative project and moral theory:

> "[V]arious camps express agreement that more careful and empirically informed work on the nature or history or function of morality is needed. Perhaps unsurprisingly, very little such work has been done even by some of those who have recommended it most firmly. Too many moral philosophers and commentators on moral philosophy - we do not exempt ourselves - have been content to invent their psychology or anthropology from scratch."

Goldman on Naturalizing Metaphysics

More generally speaking, I take the next two chapters to be done in the spirit of a view recently set out by Alvin Goldman, in his paper "A Program for "Naturalizing" Metaphysics, with Application to the Ontology of Events". The main thrust of Goldman's argument there is methodological. While he does not too strongly advocate ontological conclusions about anything other than events, he is adamant in his claim that, generally speaking, evidence provided by cognitive science can and should play a bigger role in adjudicating between philosophic options available in traditional metaphysical disputes.

Goldman begins by addressing those who might think his proposal is a non-starter: "Metaphysics seeks to understand the nature of the world as it is independently of

how we think of it. The suggestion that we should study the mind to understand reality would therefore strike many metaphysicians as wrong-headed" (page 1-2, his italics). In other words, one might think that if metaphysics seeks to get *beyond* the mere appearances generated by the mind and figure out what reality is like in a mind-independent way, apart from the ways we tend to think about it, then focusing on the operations of the mind might seem immediately self-defeating. Goldman correctly begs to differ, though. He suggests that since metaphysicians are attempting to elucidate the relationship between "appearances", on the one hand, and the reality behind those appearances, on the other, it makes just as much sense to study how and why the mind, the "aggregate of organs or mechanisms of cognition" (page 2), produces those appearances.

Thought he does not put it in these terms, Goldman can be seen as arguing that cognitive science can supplement, and perhaps in some cases supplant, traditional conceptual analysis and phenomenological introspection. That is, the findings from cognitive science can play the role that armchair conceptual analysis and phenomenological introspection have played in much metaphysical debate: they can help us better characterize and understand the nature of "appearances," broadly construed.[37] Generally speaking, Goldman endorses the traditional methodology of metaphysics, roughly understood as that of comparing "appearances" with "reality", and then deriving conclusions about the relationship between the two, and thus about the status of the appearances themselves. In his own words:

---

[37] Here and in the next two chapters, I will follow Goldman's usage of the term "appearance", and use it as a very general term that captures whatever entities, judgments, concepts, percepts, and so forth, that fall on the intuitive, usually mental side of the metaphysician's comparative enterprise, opposite the ultimate reality to which they are compared.

> "[M]etaphysical inquiries usually start with default metaphysical assumptions, i.e., naïve, intuitive, or unreflective judgments. These correspond to what he [Earl Conee, whom Goldman is discussing – DK] calls "appearances." We intuitively judge that objects are colored, that people have free will, that some events cause others, that time passes (always moving in the same direction), and that some possibilities are unactualized. Metaphysical inquiry starts from such default judgments, but it is prepared to analyze or interpret them in alternative ways, or even to abandon them altogether. They are all up for critical scrutiny, of one sort or another. How should we proceed in this critical, reflective activity? To what degree should precedence, or priority, be given to our naïve metaphysical convictions?
>
> "Virtually all metaphysicians agree that our default metaphysical views are subject to philosophical refinement. If there are inconsistencies among our naïve metaphysical views, some must be abandoned. In addition, most contemporary metaphysicians would agree that science should sometimes override our naïve metaphysics. Physics might give us reason to conclude that time doesn't "pass" at all; that it has no asymmetrical directedness; or, indeed, that there is no such thing as time, only space-time. Again, physics might give us reason to abandon certain assumptions about causal relations. Most existing appeals to science in defense of metaphysical refinements (or revolutions) are appeals to <u>physical</u> science. This is understandable, given that most of metaphysics is concerned with ostensibly non-mental targets (e.g., color, causation, time, possibilia). I argue, however, that even in these sectors of metaphysics, evidence from mental science, that is, cognitive science, can and should be part of metaphysical inquiry."
>
> (Goldman 2007, page 2)

The recommended view is that rather than *simply* relying on armchair conceptual analysis and phenomenological introspection to characterize appearances, we should supplement our understanding of those appearances that make up one side of the metaphysician's comparative exercise with the rich resources of cognitive science. The methods of cognitive science provide our best, most systematic access to the workings of the "cognitive organs or mechanisms [that] play a critical role in the causal production of appearances", Goldman maintains, and so "in considering whether such metaphysical appearances should be accepted at face value or, alternatively, should be superseded

through some sort of metaphysical reflection, it obviously makes sense to be as informed as possible about how these mechanisms of cognition work" (page 2).

Goldman points to a variety of options to choose from in drawing conclusions about some given set of appearances, and maintains evidence from cognitive science might help tilt the scales in favor of one or another of these. She might conclude, after philosophic reflection, that they should be accepted at face value. Alternatively, she might conclude that our intuitive, and perhaps naïve understanding of those appearances needs to be revised. Such revisionary approaches maintain that the relevant appearances may be slightly misleading, but they capture enough of the truth about the entities in question that they our intuitive understanding of them need only be refined in one way or another. Once again, in his own words,

> "Starting with a naïve conception of a certain property, a metaphysician might suggest that the property is really different in crucial ways from the way common sense or experience represents it. The proposal does not deny the phenomenon's existence (some phenomenon worthy of the name). It merely suggests that the property's ontological status is importantly different from the way it is ordinarily represented."
>
> (Goldman 2007, page 2)

Finally and most radically, is the eliminativist option, where the metaphysician decides, after philosophic reflection, that "This or that ontological phenomenon, assumed to exist on the basis of common sense or naïve experience, might be denied any sort of existence at all" (page 2-3).

I take the work being done in the next two chapters of this dissertation to fall under the aegis of the naturalistic framework Goldman has set out for investigating metaphysical questions and ontological issues. In these chapters, disgust again makes an ideal focal point for addressing such issues for at least three reasons. First,

disgustingness is an excellent example of the types of ontological appearances that Goldman discusses (indeed, he uses it to motivate his program). Disgust can intensely flavor our experience. It has been called the most visceral of emotions. It exerts a deep, primal effect on our perception of the world, especially of those things that we experience as disgusting. Second, what types of things are experienced as disgusting is highly variable, both from individual to individual and from culture to culture. In the grip of this consideration, one might feel drawn to outright eliminativism with respect to disgustingness. The issue certainly begs for the type of metaphysical reflection, informed by the relevant cognitive science, that Goldman recommends.

Rather than embracing bald eliminativism, others hold up disgustingness a paradigm case of a property that the mind projects upon the world. At the very least, disgust can be (and has been, for instance see McDowell 1998, D'Arms & Jacobson 2003) used to clarify the very notions of projection and projectivism. It not only provides a concrete example, but one that can serve as a simplified model for investigating the relevant issues. For, it is not completely entangled with ancillary issues like normative naturalism or worries about moral nihilism that can greatly complicate metaethical discussions about realism, projectivism, eliminativism, and the like. Given both of these properties, the drastic effect it has on our experience, together with the variability in the sorts of things that induce it, disgust is apt to prompt questions like those we began with: is anything *really* disgusting? Is disgustingness a real property of some objects, or is it simply a powerful but misleading illusion, something that is projected onto the world?

Easy answers to these questions are not immediately forthcoming. Indeed, even formulating the questions in a coherent manner is a notoriously difficult challenge.

However, the third reason that disgustingness is ideal for addressing these issues is that we now have a much more detailed understanding of the psychology of disgust. In the next two chapters, we will use this understanding to shed new light on ways to frame questions of projectivism and realism.

Chapter 5: Projectivism Psychologized: A Philosophic Idea in Cognitive Scientific Clothing

5.1 Introduction

The idea that the mind actively projects something onto the external world, rather than passively reflecting everything it finds there, has a long and venerable history in philosophic discourse. As intuitively compelling as the imagery may be, however, it is obviously still quite vague. Indeed, as is often the case, the vagueness fuels the allure. While the idea has won eminent supporters both historically and more recently, others remain unconvinced that it is anything more than an empty metaphor. Opposition to projectivist-style accounts of various properties has provided common ground for a number of philosophers who would seem to agree on little else. Recent skeptical authors have advanced a pattern of argument, which I will call "the Master argument," that is alleged to show that such accounts are not just empty or wrong, but ultimately incoherent.

Despite this resistance, I remain one of those philosophers who continues to find projectivism an attractive way of locating problematic properties in nature, and reconciling a scientific picture of the world with our lived experience. I think, however, that the notion of projectivism, which makes such explicit appeal to the operation of the mind, needs to be updated. Our understanding of the mind, in the flowering of the cognitive sciences, has come a long way since the insightful armchair speculations of David Hume, the most famous historical proponent of the idea. Given the resources of cognitive science, I believe a more detailed and coherent version of projectivism, which captures the essentials of Hume's idea, can be formulated and defended.

Therein lies the aim of this paper. After briefly sketching some relevant background by showing how the notion of projectivism has been articulated in both historical and more modern context in Section 2, we will focus on formulating this skeptical line of reasoning as generically as possible in Section 3. Having laid out the Master argument, we will set it aside, and turn our attention in Section 4 to articulating the idea of projectivism in the vocabulary of modern psychology. We then return to disgust and the property of disgustingness, which is often taken to be a paradigm example of a projected property: if anything is correctly characterized by projectivism, it seems, disgustingness is.[38] By first extending the framework from Section 4, and additionally drawing on the details of my theory of the evolution and psychological apparatus that produce disgust, I argue that our resuscitated notion of projectivism captures much of what Hume and other philosophers have had in mind when they appeal to the imagery and slogans associated with the tradition. Section 6 brings things to a close by showing that the Master argument fails to get any traction on this reconstructed view, and by responding to obvious objections to our psychologized projectivism.

5.2 The Enduring Allure of Projectivism

Projectivism, intuitively speaking, is the idea that in some cases, what we unreflectively take to be features of the world are actually features of our own minds, projected outward onto the world. A projectivist about beauty would claim that the legendary loveliness of Helen of Troy, for instance, was in the eyes of her many adoring beholders, rather than Helen herself. As is all too often the case with philosophic doctrines, however, the intuitions and imagery associated with projectivism are easier to

[38] See, for instance, McDowell 1998, page 151 agreeing with J.L. Mackie.

grasp than any detailed articulation of the idea. In contrast to the image of the mind as a mirror that passively reflects the external world to which it is held up, projectivism likens the mind to a lamp that projects light outward, and thus adding to the world it is illuminating.[39] This later image is supposed to capture the thought that in perceiving and thinking about the world, our minds slyly add something to it that was not there otherwise. We are apt to mistake those added features, elements of our own perception and thought, for actual features of the world onto which they are being projected.

In contrast to full-fledged Berkeleyan idealism, projectivism is usually not advanced as a global doctrine about the fundamental nature of reality *en toto*. Projectivist-style accounts, rather, are usually given for a particular target domain, some circumscribed phenomenon or set of appearances that are being contrasted with more objective or mind-independent features of the world. The projectivist is suggesting, only about those former features, that they are more rooted in the functioning of our minds than the more objective features of the world that our minds are in contact with. Man is not the measure of *all* things, claims the projectivist, but only those phenomena in the target domain that are being separated out and given a projectivist-style treatment.

What unites different variations of this projectivist tradition is more that they take this collection of compelling slogans and suggestive metaphors as their starting point, rather than that they share commitment to any set of precise theses or refined body of doctrine. Those slogans and metaphors have been elaborated in different ways over the course of time, however.

5.2.1 Historical Roots of the Tradition

---

[39] See, for instance, Rorty 1979; cf. Abrams 1971.

As noted earlier, a perennial concern of philosophers is how our intuitive understanding of the world is related to its actual, fundamental nature. Philosophers of radically different orientations and outlooks have addressed issues centering on whether the world is as it appears to us, and how much of what appears is illusory or misleading. Not only do such questions have a venerable pedigree, but many great philosophers are closely associated with the vivid ways they motivate these sorts of problems, and the type of answers they give to them: Plato's allegory of the cave and theory of the forms, Descartes' evil deceiving genius and subsequent appeal to God, Locke's tabula rasa and camera obscura, and his distinction between primary and secondary properties.

The vision of a mind actively projecting properties onto the world is another metaphor used to motive and address such issues. Historically, the philosopher most commonly associated with the imagery of projectivism was David Hume, whose thoroughgoing empiricism and resulting skepticism lead him to endorse projectivist accounts of a wide range of phenomena, including color, as well as aesthetic and moral value. In one of the most infamous conclusions he draws from the application of his strict empiricist principles, he declares that we project a relation of necessary connection onto the world when we observe the constant conjunction of two objects or events. We merely mistake that relation as causal; for Hume, even causation is projected.

Some of Hume's most memorable turns of phrase are expressions of his projectivist positions on similar issues. He makes the general observation that the mind has a "propensity to spread itself on external objects" (Hume 1978, 1.3.14), and speaks of the mind "gilding or staining...natural objects with colors, borrowed from internal

sentiment" (Hume 1975, p. 294). In defending his projectivism about moral values, he asks his reader to reflect upon the dim judgment she has made of some action:

> "Examine it in all lights and see if you can find that matter of fact...which you call vice...The vice entirely escapes you, as long as you consider the object. You can never find it until you turn your reflection into your own breast, and find a sentiment of disapprobation, which arises in you, toward that action."
> (Hume, 1978, pp. 468-9).

Likewise, he locates aesthetic value in the mind, rather than in what it beholds:

> "Euclid has fully explained all the qualities of the circle; but has not in any proposition said a word of its beauty. The reason is evident. The beauty is not a quality of the circle...It is only the effect, which that figure produces upon the mind, whose peculiar fabric or structure renders it susceptible of such sentiments. In vain would you look for it in the circle, or seek it, either by your own senses or by mathematical reasonings, in all the properties of that figure."
> (Hume, 1975, pp. 291-2)

Neither Hume's considerable stature nor eloquent endorsement of multiple projectivist theses can completely account for the continuing appeal of the idea, though. Rather, projectivism had remained philosophically attractive because it seems to offer a way of "locating" certain properties in the nature world, especially with respect to the natural world as revealed by modern science. Since Hume's day, the continued success and resulting epistemic authority earned by science has seemed to some to legitimate a picture of the world that Hilary Putnam (1990, 1999) has called the "World Machine." With this rise of science, it has been increasingly difficult to see where properties such as color or beauty fit into the causal order of that world, or where moral values can be located in nature – at least with respect to the "World Machine" picture of nature science seems to deliver.

Projectivism has been taken by many philosophers as a strategy for dealing with this difficulty: properties in some target domain are difficult to locate in the workings of

the World Machine because they are, strictly speaking, not part of it. Rather, they are projected onto causal machinery of the world by our minds, in much the same way that a film projector projects images and colors onto a movie screen that is actually devoid of colors and images. The projectivist claims that those problematic phenomena are not properly ascribed to the acts and entities that actually make up the causal machinery of the world, but are added, projected onto those acts and entities by the perceptual and psychological apparatus that generates our experience of them. Hume notes that the mind, in "gilding or staining all natural objects with the colours, borrowed from internal sentiment, raises, in a manner, a new creation" (Hume 1975).

5.2.2 Modern Incarnations

While Hume might be the most famous proponent of the tradition, no one owns the notion of projectivism. In 20th century analytic philosophy, dominated by the so-called linguistic turn, projectivist ideas have become associated and sometimes deeply intertwined with semantic theses about the meanings of individual terms, or the role and significance of truth conditions and the truth predicate. Since many of the questions of philosophy were reframed as questions about language, many of the slogans at the heart of this tradition were dressed up in linguistic clothing. These proposals can usually be separated out into a semantic component and a projectivist component.

As Rachels (2002, chapter 3) points out, in the wake of the linguistic turn, one early attempt to preserve Hume's insights about a projecting mind was the semantic thesis of emotivism (see Ayer 1936, Stevenson 1937). Roughly speaking, an emotivist semantics holds that the claims falling within its domain do not state facts, and are thus not the sorts of claims that admit to being true or false. Rather, despite their sometimes

misleading surface syntax, claims that are properly understood along emotivist lines merely serve to *express* the sentiments or emotions of the speaker (indeed, "expressivism" is sometimes used interchangeably with "emotivism"). In expressing a sentiment, a speaker is not asserting its existence ("I am experiencing happiness," or "Michael is angry"), ascribing some other property to it ("My happiness is boundless," or "Michael's anger is due to the traffic jam"), or making any other type of truth apt claim. Rather, the speaker is simply emoting, albeit verbally.

Authors like Stevenson also sought to extend emotivist semantics from obviously expressive turns of phrase such as "ouch!" or "hooray!" to claims about value or the moral permissibility of various acts and social practices. On this account, a claim like "abortion is morally wrong" does not, truly or falsely, predicate any properties of the practice of abortion. Instead, it serves to express the speaker's attitude of disapprobation towards the practice. In caricature, it would translate to something along the lines of "Boo! Abortion". In the hands of emotivists, then, the original, mental activity of projection is adapted to a linguistic context, where it becomes expression; the semantics of the language is supposed to be very similar to what the mind is alleged to be doing – though there is little talk of the mind itself. On this account, the semantic component and the projectivist component are indeed deeply intertwined; in fact, they are fused together.

Projectivist ideas also appear in the work of error theorists like J.L. Mackie (1977). Contrary to emotivists, error theorists hold that claims about value and morality do indeed attempt to state facts. The semantic component of this view takes the surface grammar of such claims at face value, as purporting to be about objective, mind independent entities. The claims can thus be true or false in much the same way that

scientific claims or more mundane statements about middle sized objects can be true or false. However, error theorists hold that the objective, mind independent facts that claims about value and morality purport to refer to and describe simply don't exist. Therefore, all such statements are false; the entire discourse is radically in error.

On this picture, the idea of a projecting mind comes into play to explain how we could be so radically and systematically wrong, how the entirety of our moral discourse could be founded on such an egregious error. According to such an error theoretic view, our more trusted epistemic sources have convinced us that, as a matter of contingent fact, the world we inhabit does not contain any objective, mind independent values. We make the mistake of thinking that it does, however, because our minds gild and stain that world with value-like projections in a way that makes them appear to be objective and mind independent. Our lived experience presents the world as-if it contained such values, though we have come in our wisdom to see that it does not. The semantic component of this view takes the language of value and morality as standard fact stating discourse. But it sees that discourse as purporting to refer to and describe a domain of facts that simply don't exist. The projective component of the view bears the explanatory burden of showing why we are in the grip of the illusion that they do.

Simon Blackburn's quasi-realism about value and moral discourse also draws on projectivist ideas. Blackburn's treatment (1984, 1993) is perhaps the most careful and explicit when it comes to separating out the semantic component and projectivist components. Blackburn is dissatisfied with error theories, and seeks to philosophically vindicate most of our ordinary and common sense moral thought. Doing so leads him to spend the majority of his time in the trenches of semantic debates about the character of

our moral and evaluative discourse, arguing about the behavior of the truth predicate and exploring the mysteries of the Frege-Geach problem (see Geach 1965, Blackburn 1984, chapter 6). While he nominally acknowledges the projected character of much of morality, his primary goal is to show that the evaluative discourse we use to talk about it is robust and well behaved enough to be captured by classical logic and to sustain the proper use of a truth predicate (albeit one understood along minimalist or deflationary lines). Thus, "quasi-realism" is the name of his interpretation of the linguistic character of that *discourse*, rather than a thesis that directly characterizes the metaphysics of value or psychology of projection. The linguistic work done by the quasi-realist semantic interpretation then shields the discourse and its domain of discourse from charges of irrealism or radical error. In his own words, "projectivism can accommodate the propositional grammar of ethics. I need not seek to revise that. On the contrary, properly *protected by quasi-realism* it supports and indeed explains this much of our ordinary moral thought" (1993, page 153, my italics).

Error theorists like Mackie and quasi-realists like Blackburn are united in rejecting emotivism, and instead maintain that moral claims purport to state facts, have truth conditions, and can thus be true or false. While error theorists hold such claims are uniformly false, and invoke the projective character of the mind to explain why it seems otherwise, quasi-realists hold that at least some such claims are true, and invoke the mind's projective character and the properties that it projects to explain what those claims might be true *of*.

This is by no means an exhaustive survey. Rather, the positions mentioned can be thought of as emblematic of ways that projectivist slogans have been fleshed out in the

context of more recent philosophic debates. Since it is more a philosophic term of art, or better a tradition or a school of thought, no one will have the last word on projectivism. Curiously, though, while all of these views reserve a crucial role for the projecting mind to play, none of them have much to offer by way of what it might mean for the mind to project anything onto the world. Those dubious of the tradition take them to task for that omission.

5.3 Resistance: The Master Argument Against Projectivism

As mentioned above, projectivist style accounts are local, applying to some circumscribed phenomenon, such as causation, or a delimited type of property, such as color. While they are always local, they can be given for a number of different types of properties. In other words, every projectivist account has what I will call a target domain. As we will see, when critics of projectivist style accounts mount their criticism, they often do so against some particular projectivist account, i.e. projectivism about color, or projectivism about value, etc. Indeed, the critics attack many accounts that have a projectivist *flavor*.[40] These arguments, however, often share an underlying form, logic and conclusion, namely that those projectivist style accounts and the uses to which they are put are ultimately incoherent. I will call such lines of criticism versions of the Master argument.

5.3.1 Variations on a Theme

---

[40] Terminology is a problem here, as different authors give the same terms slightly different nuances, or use different jargon to characterize views that have much in common. I am interested in the idea of projectivism, and have sought to construe it rather broadly, at least to start. What I'm calling the Master argument, however, has been advanced against views called projectivist, but also views called response dependent, dispositionalist, and even some views that have been called eliminativist. As I'm simply calling them all projectivist *style* accounts for now. The point I wish to highlight in doing so is that the Master argument, whatever particular view it is used to attack, always shares the same underlying logic, and contains the same set of moving parts.

The first version we will look at is due to Stephen White, who advances an elegant but compressed version of the Master argument. In it, however, we will see many themes that will become familiar components of the line of attack: a foil, who endorses or is saddled with some restrictive metaphysical or ontological view. Certain types of properties are acknowledged as fitting uncomfortably within the restrictions of that metaphysical view, and so the imagery associated with projectivism is invoked to account for them, or explain them away. In White's case, the foil is Galen Strawson, who subscribes to an objectivist metaphysics. In construing agency and free will as illusions, Strawson advances an account of them that bears many of the characteristic marks of projectivism:

> "Strawson suggests that freedom is really a kind of necessary illusion but fails to ask the genuinely deep question – namely, what the illusion is an illusion *of*. If freedom and agency are incoherent on the assumption of determinism and, equally, are incoherent on the assumption of randomness and on any other assumption about the objective metaphysical facts, then from an objectivist metaphysical perspective the notion is incoherent. It is then a mystery *what* people *think* they have when they think they have free will. … Strawson provides no account of the *content* of the illusion of freedom, and prima facie both the idea that the future is completely fixed and the idea that there are random events make the idea of action *incoherent*."
>
> White 2004, page 203-4, his italics

Barry Stroud develops another version. Stroud's foil is a naturalist who defends some form of dispositionalism about color. He again sounds many of the main themes of the Master argument, and is eloquent and thorough enough that I will let him speak for himself:

> "What human beings think, feel, and care about must be fully expressible somehow with the restricted resources available in the naturalist's world. And that can lead to distortion. If, to accommodate psychological phenomena and their contents in all their complexity, the restrictions are lifted, naturalism to that extent loses its bite. This is the basic dilemma I want to bring out. … For example, many philosophers now hold that things as they are in the world of

> nature are not really colored. There are rectangular tables in the natural world, perhaps, and there are apples in the natural world, but no red apples (and no yellow or green ones either). This view appears to be held largely on the grounds that colors are not part of "the causal order of the world" or do not figure essentially in any purely scientific account of with is so. Scientific naturalism accordingly excludes them. … But even on this view those false beliefs and illusory perceptions of the colors of things must themselves be acknowledged as part of nature. A naturalistic investigation must somehow make sense of them as the psychological phenomena they are. Since he holds that there is no such fact as an object's being colored, he cannot specify the contents of those perceptions and beliefs in terms of any conditions that he believes actually hold in the world. If he could, that would amount to believing that there are colored things in the world after all. … A dispositionalist theory … can succeed only if it can specify the contents of the perceptions of color, which it says physical objects have disposition to produce. They cannot be identified as perceptions of an object's having disposition to produce just *these* perceptions under certain circumstances. The question is: Which perceptions? There must be some way of identifying the perceptions independently of the object's disposition to produce them. So it looks as if they must be identified only in terms of some so-called "intrinsic" quality that they have. Not a quality that the perception is a perception *of*, but simply a quality of the perception itself. … I doubt that we can make the right kind of sense of perceptions of color in this way. So I doubt that any dispositional theory can give a correct account of the contents of our beliefs about the colors of things. The way we do it in real life, I believe, is to identify the contents of perceptions of color by means of the colors of objects they are typically perceptions of. It is only because we can make intelligible nondispositional ascriptions of colors to objects that we can acknowledge and identify perceptions as perceptions of this or that color. But if that is so, it requires our accepting the fact that objects in the world are colored, and that is what the restrictive naturalist who denies the reality or the objectivity of colors cannot do. … Most philosophers regard it as so obvious and uncontroversial that colors are not real, or are in some way only "subjective," that they simply do not recognize what I think is the distortion or incoherence they are committed to."
>
> Stroud 1996, page 27-9

It is clear that he thinks the Master argument applies equally well to projectivist-style accounts of value as well as color. He states that "To understand and acknowledge the presence of these human [evaluative] attitudes in the world, the naturalist must understand their contents – what those human beings actually think or believe" (page 29). He goes on:

> "If such a reduction is expressed in terms of the dispositions natural objects or states of affairs have to produce certain reactions in human beings, it faces the same kind of problem as the dispositionalist view of colors. Those reactions themselves must somehow be identified, and if they are left as reactions with evaluative contents, no naturalistic progress will have been made."
>
> Stroud 1996, page 30

Hillary Putnam takes on the idea of projectivism in all of its forms, but in initially formulating his argument takes projectivism about color to be his target. He first takes projectivism to be captured by

> "The idea that there is a property all red objects have in common – the same in all cases – and another property all green objects have in common – the same in all cases – is a kind of illusion, on the view we have come more and more to take for granted since the age of Descartes and Locke."
>
> Putnam 1999, page 592

Whereas Stroud's foil is the naturalist, Putnam's is what he calls the Objectivist, who is committed to:

> "the 'fundamental Objectivist assumptions', … 1) the assumption that there is a clear distinction to be drawn between the properties things have 'in themselves' and the properties which are 'projected by us' and 2) the assumption that the fundamental science – in the singular, since only physics has that status today – tells us what properties things have 'in themselves'. … So to explain the features of the commonsense world, including color, solidity, causality … in terms of a mental operation called 'projection' is to explain just about every feature of the commonsense world in terms of *thought*. The problem, in a nutshell, is that thought itself has come to be treated more and more as a 'projection' by the philosophy that traces its pedigree to the seventeenth century."
>
> Putnam 1999, page 595

But given the Objectivists commitments, she "will have to conclude that intentionality *too* must be a mere 'projection'." Thus we arrive at the charge of incoherence: "But how can any philosopher think this suggestion has even the semblance of making sense? As we saw, the very notion of 'projection' *presupposes* intentionality!" (Putnam 1999, page 596). He ends by portraying the incoherence as stemming from a clash of conflicting conceptual schemes:

> "If this is right, then it may be possible to see how it can be that what is in one sense the 'same' world (the two versions are deeply related) can be described as consisting of 'tables and chairs' (and these described as colored, possessing dispositional properties, etc.) in one version *and* as consisting of space-time regions, particles and fields, etc. in other versions. To require that all of these *must* be reducible to a single version is to make the mistake of supposing that "Which are the real objects?' is a question that makes sense *independently of our choice of concepts*."
>
> Putnam 1999, page 598, his italics

John McDowell, in his "Projection and Truth in Ethics," mounts a version of the Master argument as well. He sees "The point of the image of projection is to explain certain seeming features of reality as reflections of our subjective responses to a world that really contains no such features" (McDowell 1998, page 157), and begins his attack by alluding to the problem that inevitably arises with the image: "The right explanatory test is not whether something pulls its own weight in the favoured explanation (it may fail to do so without thereby being explained away), but *whether the explainer can consistently deny its reality*" (Page 142, my italics). Speaking of the idea that fearfulness is projected onto the world, he claims:

> "So explanations of fear that manifest our capacity to understand ourselves in this region of our lives will simply not cohere with the claim that reality contains nothing in the way of fearfulness. Any such claim would *undermine the intelligibility* that the explanations confer on our responses."
>
> McDowell 1998, page 144, my italics

5.3.2 The Core of the Master Argument

In one way or another, all of these critiques question the coherence of the notion of projectivism and the uses to which it is put; whether the idea can even be made sense of is called into doubt. Versions differ in their level of clarity and each is marked by its own unique subtleties. The perspective afforded by generalizing away from the specifics of any one version, however, allows us to see that there are really three core pieces in the

machinery of the Master argument. They are 1) an overarching but constraining framework of some kind, 2) the problematic properties in the target domain or domains, and 3) the appeal to projectivism that casts those properties as illusory. In short, the puritanical imposition of a *restrictive metaphysical view* – naturalism or objectivism or some other "ism" that limits ones metaphysics to a preferred conceptual scheme and the entities it deals in – generates what I will call *renegades*. These renegades in turn give rise to the difficulties projectivism is then invoked to solve.

Once the puritan-cum-projectivist imposes her restrictive metaphysical view and some domain is taken as epistemically privileged or ontologically basic, problem areas in our intuitive or common sense ontology are then identified. Using only the resources the metaphysical puritan has restricted herself to, she needs to be able to account for the renegades that fall within these problem areas. These renegades often include colors, beauty, values and so forth, properties or appearances that seem manifest or indispensable in one way or another, but which do not look like they fit straightforwardly into the restricted metaphysical framework. Bald eliminativism not being a congenial option, the need to locate and make sense of such renegades drives the puritan to projectivism. In order to account for the problematic properties of a target domain, the puritan appeals to the imagery of projection: the renegades are difficult to account for with the concepts or within the ontology allowed by the restrictive metaphysical view because they are not really there, not really part of the austere world to which the puritan is committed. Rather, they are a sort of illusion, a projection of our own making, the gild and stain of our minds, nothing more.

It is at this last move that proponents of the Master argument cry foul. They claim that the puritan-cum-projectivist has here unwittingly hoisted herself on her own petard. The properties that require a projectivist treatment cannot be made sense of this way, cannot be accounted for within the restrictive metaphysical view, because the very notions of projection and a projecting mind cannot be made sense of within the restrictive metaphysical view, either. The projectivist resources appealed to in order to account for renegades are just as problematic as the renegades themselves. The projectivist is reaching outside the very boundaries she has set for herself. Thus the treatment for the problem is no better off than the problem itself; the cure is a bad as the disease.

Moreover, the proponents of the Master argument point out that given that the restrictions the puritan-cum-projectivist has imposed upon herself are often conceptual, she does not even have the conceptual resources to specify the content of the alleged illusion generated by the projecting mind. Given the conceptual apparatus available in the restrictive framework, she cannot even say what subjects in the grip of a projected illusion *think* they are beholding, or what they *mean* when they attempt to describe it. Thus, she is unable to even specify what she wishes to denigrate as a mere projection. According to the Master argument, then, a projectivist account is not an account of an illusion, but an illusion of an account.

5.4 Cognitive Science and Projection

> "Philosophers who talk this way rarely if ever stop to say what *projection* itself is supposed to be. Where in the scheme does the ability of the mind to 'project' anything onto anything come in?"
>
> Putnam 1999, page 594

Putnam asks a fair question: what exactly does it mean to say the mind 'projects' anything onto anything else? What *could* it mean? Can this illusion of an account be

developed into an actual one? As we saw in section 2, modern philosophers who invoke the idea often focus on the linguistic side of the issues they are interested in, spending much of their effort investigating the semantics of claims made about entities and properties that fall within the target domain. They say relatively little about the psychological side, however, apparently assuming that the details will take care of themselves. Proponents of the Master argument doubt that they will.

The aim of this section, then, is to find a way to understand what it could possibly mean to say that the mind projects some property onto the external world, instead of, say, finding it there to begin with. Rather than address the Master argument directly, however, we would be better served to take a step back. We began the paper by briefly surveying the tradition from which the idea stems, and looking at current resistance to the tradition, the Master argument whose conclusion is that projectivism is incoherent. For now, we will set both the history and the resistance aside, and instead focus on formulating a new way of understanding the idea that is firmly rooted in the cognitive sciences. It will be best to get a running start before tackling the subtleties involving disgust and disgustingness in the next section. Therefore, we use this section to sneak up on a new way of understanding a projecting mind, gradually building a vocabulary and set of tools. These we can develop by beginning with a relatively uncontroversial example.

5.4.1 Loosening Up Intuitions: Anthropomorphism

Consider the ages old human tendency to anthropomorphize. When we anthropomorphize something, we incorrectly ascribe an array of human characteristics to a thing that does not actually have them, be it a cloud, an animal, or perhaps even the

entirety of nature. The human tendency to do this is widespread, and thought to lie behind a variety of phenomena, but it is perhaps most consistently invoked in discussions of religion (see Guthrie 1993, chapter 3 for a useful overview). More specifically, the fact that we read human characteristics into non-human entities and phenomena has also been linked to the persistent belief in supernatural agents of all sorts, deities who control natural phenomena, and so forth.

Cognitive science has recently made progress in understanding this tendency of ours. More specifically, work on our folk psychological capacities (Leslie 1987, Baron-Cohen 1995, see Nichols & Stich 2004 and Goldman 2006 for overview and discussion) has been used to inform work on the psychological underpinnings of religion and religious belief (Barrett 2000, Atran 2001, Boyer 2001, Dennett 2006, especially chapter 4). This cross-pollination of ideas has turned out to be exceptionally fruitful. Our tendency to anthropomorphize has been traced to our capacity to detect the presence of other animals, especially people, to see them as animate, purposeful beings and to make sense of them in terms of their beliefs and desires. On the one hand, cognitive scientists specifically exploring our folk psychology have uncovered a number of surprising features of those folk psychological capacities and posited a variety of cognitive mechanisms that might underlie them. These include mechanisms dedicated to agency detection, which interpret certain types of motion as the volitional and purposive behavior of animate creatures, rather than the mere movement of inanimate objects. They also include mechanisms dedicated to aspects of mentalizing, which do things like ascribe intentions and mental states to those (alleged) agents, and allow easy explanation and prediction of their behavior in terms of those beliefs, desires, and other mental states.

On the other hand, cognitive scientists working on religion point out that these folk psychological capacities, together with many of their most noteworthy features, naturally explain aspects of anthropomorphism. Each of these two literatures is large in its own right, and together they are enormous. For our purposes, however, we can boil down the relevant findings to a triad of general properties that characterize the operation of the cognitive mechanisms involved, and a fourth property that characterizes *when* they operate.

Let us begin with this fourth property. Research has found that human folk psychological capacities are on a *hair trigger*: for a variety of evolutionary reasons, they follow the logic implicit in the phrase "better safe than sorry" (better to mistake a windblown leaf for a predator than mistake a predator for a windblown leaf). The underlying mechanisms are activated at the slightest provocation. Due to this, they are also apt to yield many false positives. Misfiring in such cases, they attribute agency and minds to things that are manifestly not agents, and which manifestly do not have minds, such as windblown leaves, clouds, or entire mountains. In his discussion of religion, Dennett (2006) identifies this hypertrophy of (what he calls) the intentional stance as the core of anthropomorphism and belief in supernatural agents: "At the root of human belief in gods lies an instinct on a hair trigger: the disposition to attribute *agency* – beliefs and desires and other mental states – to anything complicated that moves" (page 114).

It is not just their easy activation that is of interest, though. Equally relevant is the manner in which they do their work once they have been activated. The cognitive mechanisms underlying our folk psychological abilities, and thus our tendencies to anthropomorphize, are fairly *autonomous*, they operate *implicitly*, and they are

productive. Let us take these in order. First, the mechanisms are autonomous in that they can operate along side, and at the same time as, a variety of other parts of the mind, and while our attention is elsewhere. The operation of the mechanisms underlying our folk psychological capacities does not preclude the simultaneous operation of, for instance, the complex mechanisms involved with language production and comprehension, mechanisms subserving perception in all five modalities, mechanisms underlying higher order reflection and judgment, and so forth. Nor does the operation of these later types of mechanisms preclude our folk psychological capacities, either.

Second, to say that many of the cognitive mechanisms uncovered appear to work implicitly is to say that, for instance, mechanisms of agency detection and mentalizing often operate quickly, spontaneously and automatically, without any deliberate effort or purpose on the part of the subject. One does not simply decide to turn them on, nor can one decide to turn them off, either (though on some occasions one may be able to suppress or override their effects). The mechanisms often operate without our explicit awareness; we simply and naturally think of other people, and any other targets of our mentalizing abilities, as the possessors of minds. We just as effortlessly make the complicated inferences about the connections presumed to hold between their movements and those mental states as well. Again, the mechanisms are autonomous enough, and all of this happens so automatically and naturally, that we often do not even notice that we are doing it, let alone notice the complexity of the inferences we are performing and the scope of the assumptions we are implicitly making.[41] Moreover, such autonomous and

[41] To experience your own folk psychological capacities in action, and in a way that illustrates many of these properties, consult the variation of the famous Heider and Simmel films at this website: http://cogweb.ucla.edu/Discourse/Narrative/heider-simmel-demo.swf.

automated mechanisms can continue to operate despite cutting against our more considered judgments.[42]

Third, the cognitive mechanisms that underlie our folk psychological capacities are *productive*. Once activated by the prototypical types of motion that trigger them, these mechanisms go on to infer the presence of a wide variety of other attributes associated with agency and minds. Based on the detection of fairly limited or specific evidence, they produce a relatively large set of cognitive effects, including assumptions about other features possessed by the detected triggering entity, expectations about how it will behave, and typical patterns of inference about how best to think about and deal with it. Or in more colloquial terms, with productive cognitive mechanisms, you get more *out* than you put *in*.

To render the idea of productivity more picturesquely, think of some behavioral capacity as being subserved by a black box cognitive mechanism, a machine that takes inputs and delivers outputs. When the machine receives an input, perhaps via detection of a particular property in the surrounding environment, it performs its proprietary computations, and delivers its output. To say that the machine or mechanism is productive is to make a claim about the character of the output, namely that it is multifaceted, and consists of not a single effect but many, an entire *cluster* of them (cognitive, behavioral, affective, or otherwise). For instance, a productive mechanism

---

[42] Other well known examples is this sort of phenomenon, where automated cognitive mechanisms produce effects and appearances that cut against our more considered judgments, involve the persistence of certain optical illusions, such as the muller-lyer illusion. Such illusions remain even when subjects reflectively know that their perceptions are misleading or illusory. See Fodor (1983) for an extended discussion. He calls cognitive mechanisms that operate so persistently "cognitively impenetrable": their proprietary processes and information database are not influenced by or accessible to the central systems that underlie reflection and the like. For other examples of implicit mechanisms diverging with considered judgments, see the extensive literature on implicit biases (Banaji 2001, Greenwald et al. 2003, see also Kelly et al. forthcoming for overview and discussion of specifically racial biases).

might go from an input of detected evidence and specific environmental cues to an output consisting of a rich set of assumptions, expectations, and inferences about those inputs. In the case of our folk psychology, input triggers include things like specific types of movement, bilaterally symmetrical patterns, and perhaps language-like sounds. The output includes not only the automatic ascription of agency and mental states to the triggering entities, but an entire cluster of assumptions about the way the (putative) beliefs, desires, and goals relate to each other and expectations about the types of behavior these will give rise to in the (putative) agent. This complicated but patterned set of expectations and assumptions far outstrips what has been, or often can be, known about the triggering entity based solely on the input, the preliminary evidence that was initially detected.

Much of the research on and debate about those folk psychological capacities focuses on the character of those mechanisms inside the black box – whether they are best understood as simulation or theory based, whether they are learned or innate, the degree to which they are modularized, and so forth. What is no longer seriously doubted is that many of those mechanisms are fairly autonomous and productive, that they are on a hair trigger, and that they operate implicitly. The emerging consensus in research on the psychology of religion is that at the heart of belief in gods and other supernatural agents is a robust anthropomorphism, and that the human tendency to anthropomorphize what it finds in the world can be explained by appeal to these features of our folk psychological capacities.

5.4.2 From Autonomous, Implicit and Productive Mechanisms to A Projecting Mind

We have boiled down a large number of the features of cognitive mechanisms underlying our folk psychological capacities to just four properties (albeit high level ones), mainly for easy of exposition: implicit and autonomous operation, hair trigger activation, and productive output. These terms describe a complex cluster of mechanisms. They also wear on their sleeve that they describe the functioning of the anthropomorphizer's *mind*, since they characterize the mental operations that give rise to her anthropomorphic tendencies. Indeed, as we shall see, psychologists working in other domains of cognition have posited mechanisms that share many of the same properties.

This way of talking about how the mind works is also couched in a highly theoretic vocabulary that was developed in conjunction with controlled, scientific experiments, and is employed to understand and characterize the functioning of the mind in maximally objective, mechanistic, third person terms. But there is another way to talk about these very same features of the mind, one that better captures how the world is presented to the awareness of a person when they are in the grip of experiences shaped by their own cognitive mechanisms of this sort.

Consider again the case of anthropomorphism. On the one hand, we can explain this tendency as we have above, in that maximally objective, third person, theoretic vocabulary. In these cases, what cognitive science has found is happening is that the mechanism (or interlocking set of cognitive mechanisms), which is on a hair trigger, is activated by some non-human or inanimate feature of the environment. The mechanism then operates implicitly, and, due to its productive character, gives rise to a cluster of inferences, expectations and assumptions that are so rich as to far outstrip what has been observed about the triggering object. In the case of anthropomorphism, many of those

inferences, expectations, and assumptions are misplaced: the more specific inferences and expectations may be disappointed exactly because the inanimate triggering entity simply does not possess the agency or mental states that the cognitive mechanisms automatically attribute to them.

For the uninitiated, though, this can be hard to get one's head around, or at least the jargon can be hard to penetrate. Another, much more intuitive way to describe what happens is in terms of projection. In fact, top researchers on religion very easily fall into this type of language. In making one argument, Pascal Boyer does so a number of times:

> "We **project** human features onto nonhuman aspects of the world … [we] do not always **project** onto these agents other human characteristics, such as having a body, eating food, living with a family or gradually getting older. Indeed, anthropologists know that the *only* feature of humans that is *always* **projected** onto supernatural beings is the mind"
>
> (Boyer, 2001, page 143-4, his italic, my bold).

From the first person point of view, people are often not aware of the operation of these cognitive mechanisms; they do their work implicitly. By and large, we do not have to initiate, consciously monitor or effortfully guide them as they perform their functions. It is hardly noticeable, therefore, that the activity of the mind is responsible for the attribution of those features, and it can easily seem as if they were "out there" in the world to begin with. This effect is enhanced by the fact that the relevant mechanisms are autonomous enough to be operating at the same time that a person's conscious attention is elsewhere. Thus, the accompanying experience is simply presented in such a way that the entities in question seem to have agency and minds: they are automatically treated, and thought about, *as-if* they have those properties, whether or not they do. Perhaps that is putting the cart before the horse, though, and it is best described the other way around: because those autonomous, productive mechanisms implicitly induce such a rich variety

of expectations, assumptions, and inferences about what triggers them, the associated appearances implicitly present the entities in question as-if they actually did have the human properties associated with agency and minds.[43] Because these workings of the mind are so automatic and effortless, it is easy to see how an unreflective person might mistakenly take the source of those expectations and assumptions to be in some feature of the triggered entity itself, that was detected "out there" in the world, rather than being implicitly generated by a productive, autonomous component of her own mind that is shaping the experience.

Despite how compelling these subjective appearances may be, with a little reflection it is easy to wean oneself from taking a naïvely realist stance towards them in the cases of blatant anthropomorphism. (We will consider how to treat less obvious examples below). Once this is achieved, it is often easier to switch to a different way of talking about the experience, to a vocabulary that is not as baldly mechanistic as the jargon of cognitive science, but one that nevertheless explicitly marks the role of the anthropomorphizer's mind in producing those appearances. Indeed, it seems much easier to say that the mind *projects* agency and mentality onto the entities it anthropomorphizes, and then treats them accordingly. Because agency and mentality have been projected onto those entities, we treat them as-if they did, indeed, have the features of volitional movement, and were driven by beliefs, desires and goals. Our tendency to interact with them in certain ways, to make certain inferences about them, or to have the types of

[43] Though I'm not arguing for it here, I see no reason such clusters of effects could not influence and manifest in the perceptual phenomenology associated with the experience. For instance, see White (2004) and Noe (2004) for accounts of perception that are far richer than can be accommodated by the usual, merely pictorial metaphors.

assumptions and expectations we typically do have, is merely encapsulated in the shorthand of projection talk.

There is nothing pernicious in this kind of talk; it can easily be unpacked in terms of the existence and properties of those mechanisms uncovered by cognitive science that it seeks to gesture at but gloss over. Moreover, talk of projection strikes a middle ground between a purely, and perhaps overly credulous first person point of view that takes the appearances at face value, on the one hand, and highly theoretic, strictly third person talk about the operation of the mechanisms which give rise to that experience, on the other. In one fell swoop, it is able to countenance both the realistic flavor of the experience, the fact that it seems as though the ascribed properties are actually "out there" in the world, inhering in the entities they are projected onto, as well as the knowledge of the fact that their source is actually in the operation of the components of the mind that give rise to the experience itself, and that the appearances of specific properties are generated by the way the mind cognizes entities that fall within a particular domain.

We have focused on folk psychology and religion here, but mechanisms bearing these properties might not be particularly rare, if the last few decades of research in cognitive science is any indication. Indeed, it appears that at least some capacities underlying, for instance, racial cognition operate implicitly, and are productive as well. Recent research has shown that racial biases generated by implicit cognitive mechanisms often coexist with non-racial attitudes professed by subjects when explicitly asked. Moreover, from sensory properties such as skin color, people are apt to project racial "essences" onto all members of a race. The mind's propensity to project these racial "essences" is thought to explain widely held, but largely false, beliefs about physical,

behavioral and moral properties that are taken to be characteristic of a race, and present in all of its members (see Kelly et al forthcoming). Once again, we can talk of the projection of "essences", or we can unpack that talk in terms of the rich set of inferences, assumptions, and expectations that are automatically generated by the productive and implicit mechanisms that underlie such racial cognition. (The same might be said of other instances of this type of "essentializing," as in folk biological judgments; see Gelman 2003).

While the talk of projection is not always present or foregrounded, other cognitive scientists have argued that indeed, much of the mind does not fit the intuitive, Cartesian picture of mental operations that are easily available to introspective access, and that are under direct conscious control. Such aspects of mentality are coming to look like a smaller and smaller portion of the rich, variegated tapestry of mental life, which is dominated instead by highly autonomous, implicitly operating, productive mechanisms. When it comes to the projective character of our own minds, we are too often "strangers to ourselves" (Wilson 2002).

5.5 Projecting Disgustingness

Much has been previously said about disgust and the cognitive mechanisms that underlie the emotion, and we need not rehash it here. Luckily, the work done in the last section can be extended fairly straightforwardly to disgust, and used to shed light on the common suspicion that disgustingness is one property that is always projected onto the world.

5.5.1 Disgustingness and the Disgust Execution Subsystem

The theory of disgust developed in Part I of this dissertation revealed a set of cognitive mechanisms that shared many of the relevant properties with those underlying our folk psychological capacities. Following a similar evolutionary logic of "better safe than sorry" the disgust execution subsystem is also on a hair trigger, and this also gives rise to a well-documented set of fairly straightforward false positives.

The mechanisms underlying disgust are productive as well. The detection of a certain type of food, the smell of a rotting corpse, or the violation of a purity norm will activate the entire nomological cluster, the full suite of components that make up the disgust response. These include a gape face, a flash of nausea and sense of oral incorporation, as well as a quick withdrawal and more sustained sense of offensiveness and contamination sensitivity. These constitute a set of expectations, inferences, a`nd assumptions similar to those discussed earlier in relation to anthropomorphism and our folk psychological capacities. In the case of disgust, the triggering object is assumed to be aversive, expected to be harmful. Patterns of inference are made about the disgusting object, including thinking of it as dirty or tainted, and about its ability to transmit its disgustingness to other entities that it comes into contact with. And while the productive output of the disgust mechanisms may not be as *cognitively* complex as the inferences about mental state and their relations to behavior generated by the agency detection and mentalizing mechanisms, the productivity of the disgust execution subsystem is more diverse. That is, it does contain some cognitive components, but being an emotion, it contains other types of elements as well. Activation of disgust produces characteristic behavioral components like the quick withdrawal and gape face, and characteristic affective and physiological components, such as nausea and a slight dip in heartbeat.

While there is often a hard (but not impossible) to miss phenomenological component of disgust, there are many other components to the response, and the cognitive mechanisms operate implicitly to produce these. The response is reflex-like in that it can be effortlessly and automatically triggered, and the patterns of inference associated with contamination sensitivity and offensiveness can seem entirely natural. Surprisingly, the relevant mechanisms can be triggered without our awareness, and even influence our higher-level judgments having to do with moral permissibility without our being aware of their involvement, as demonstrated by Wheatley & Haidt (2005). The expectations, inferences, and intuitive judgments they produce or influence can cut against our reflective judgment in much the same way, and can do so just as persistently as other implicit mechanisms, as illustrated by many people's reluctance to eat a turd-shaped chocolate or drink juice from a new, sterile bedpan or stirred with a new, unused comb (see Rozin et al 2000 for an overview).

As in the case of anthropomorphism, we may think about disgust and disgustingness in terms of projection. Once again, it is often easier to switch to a vocabulary that is not as baldly mechanistic as the vocabulary of cognitive science, to one that nevertheless explicitly marks the role of the disgusted subject's mind in producing the experience of disgust and the "appearance" – in Goldman's (2007) sense – of disgustingness. Indeed, it is quite natural to say that the mind *projects* disgustingness onto the entities that trigger disgust. This is true for the same reasons that it seemed easier to say that in cases of anthropomorphism, the mind projects agency and mental states onto entities in the world, rather than passively reflecting what it finds there. Since we do not have to initiate, consciously monitor or effortfully guide the relatively

autonomous mechanisms that produce the emotion of disgust, it seems like – the experience is simply presented in such a way that – the entities in question actually are offensive, tainted, and contaminating: they are automatically treated, and thought about, *as-if* they have those properties. Or, again, perhaps the puts the cart before the horse, and it is best described the other way around: because those productive mechanisms implicitly induce such a rich variety of expectations, assumptions, and inferences about whatever triggers them, the appearance of disgustingness, indeed, the very perceptual experience correlated with their operation presents the entities in question as-if they actually were bad, nauseating, tainted, and contaminating. Our tendency to make those inferences, or to have the types of assumptions and expectations we typically do have, is captured by the shorthand of projection talk – we project the property of disgustingness onto entities in the world, and then treat them accordingly. The property of disgustingness, which seems like it is a property of things "out there" in the world, is in fact an encapsulation of the suite of components of the disgust *response* to the things that trigger these implicit and productive mechanisms. Saying we project the property of disgustingness is just saying that we naturally treat such entities as-if they were offensive, tainted, contaminating, and so on.[44]

5.5.2 Imperfect Fit and the Pragmatics of Projectivist Explanations

---

[44] As we noted when we first characterized the disgust response in the behavioral profile:

> "It is also worth emphasizing that though "offensive" and "contaminating" are properties often ascribed to *items* that trigger disgust, a sense of offensiveness and contamination sensitivity and the patterns of behavior associated with them, in the sense discussed here, are parts of the *response* to such items. Indeed, one of the most insidious aspects of disgust is that once an item triggers it, that item is thereby treated *as if* it were offensive and contaminating – whether or not it is genuinely offensive (if there is such a thing) or objectively contaminating (which there certainly is). In this sense, then, it is part of the disgust response that the properties of offensiveness and contamination potency are *projected onto* whatever elicits it." (Chapter 1, page 7).

One question that arises is whether the idea of projection is appropriate when the set of inferences, expectations and assumptions encapsulated by the talk of projection is largely correct. It is natural to talk about our tendency to anthropomorphize in terms of projection, but what about when our folk psychological capacities are activated by, say, another person, who is animate, who does possess agency and have a mind, and whose behaviors are connected to their beliefs, desires, and other mental states in just the agency detection and mentalizing mechanisms cognize them? Does the mind project *only* in cases where the mechanisms involved are yielding a false positive?

Given the reconstructed the notion of projection we are working with, the answer to this question is straightforwardly "no". For, given our reworked understanding of projection in terms of productive, autonomous cognitive mechanisms and their implicit operation, it becomes clear that the mind is "projecting" whenever those mechanisms are activated, whether the triggering entity is an anthropomorphized cloud or a fully animate human being, complete with beliefs and desires.[45]

However, it is certainly more natural to talk in terms of projection in cases of anthropomorphism because there is an obvious need to appeal to the role of the projecting mind, namely to explain the fairly obvious error involved in anthropomorphism. When the mechanisms of the mind, and cluster of assumptions, expectations, and inferences they produce, are more seamlessly fitted to their object, that need does not arise. Since

---

[45] In the case of folk psychological capacities and the mental states that they ascribe, it is interesting to note that while there is currently a large if loose consensus regarding the existence of some type of dedicated cognitive mechanisms underlying our ability to mentalize, there have been genuine philosophic debates about whether those mechanisms could yield anything *but* false positives. Philosophy of mind in the 80's was dominated by debates over realism and eliminativism about common sense mental states, which explicitly addressed the question of whether the beliefs, desires, and other mental entities ascribed by those folk psychological capacities exist *at all*, in humans or anything else (Dennett 1981, Churchland 1981, Stich 1983, Fodor 1987). What no one challenged, however, was that we do, indeed, ascribe, make sense of and predict each other in terms of such of mental states.

there is less explanatory work to do, talk of projection, of the mediation of the relevant cognitive mechanisms, becomes otiose from the point of view of the pragmatics of explanation. Thus, projection talk often simply drops out. But, from the absence of that explanatory role and disappearance of the pragmatic need to fill it by appeal to the projecting components of the mind, it does not follow that the mind itself is not projecting.[46]

The situation is to some extent similar when we move from our folk psychological capacities to the mechanisms underlying production of disgust and the analogous questions that arise about disgustingness. Does the mind *only* project disgustingness in those cases where the mechanisms involved are yielding a false positive? Again, given the notion of projection we are working with, the answer to this question is straightforwardly "no".

But in the case of disgust, the situation is importantly and interestingly different as well. For, given what we know of the disgust response, the cluster of components that constitute it, and the various roles that it has been co-opted to play, we know that there are, coarsely put, almost no true positives at all. For a variety of reasons, there is nearly always an *imperfect fit* between the full disgust response and the entities that trigger it. Vague awareness of this imperfect fit, in turn, can raise suspicions that something is amiss. Careful reflection can refine that suspicion, and thus create an explanatory role that projectivist talk can be very useful in filling. Since, pragmatically speaking, imperfect fit creates the felt need for further explanation, and there is nearly always an

---

[46] To put the point in terms of one of the images at the heart of the projection, a film projector could easily project a blue image onto a screen that is, itself, the same shade of blue as the image projected onto it. Two conclusions can be drawn from this possibility. Knowing that a blue image is being projected, one is not thereby licensed to infer that the screen is *not* blue. Alternatively, knowing that the screen is blue, one is not thereby licensed to infer that nothing blue is being projected onto it, either.

imperfect fit between response and triggering entity in cases of disgust, there is nearly *always* an explanatory role that the appeal to the projecting mind can fill.[47]

Imperfect fit can come in degrees. The most flagrant cases are the clear false positives. With disgust, these are often generated by the hair trigger of the response by things like turd shaped chocolates and juice stirred with a sterilized cockroach. Here the triggering entities are obviously neither poisonous nor infectious nor contaminating, and thus are not matched to any of the individual components of the response they elicit. Imperfect fit can take subtler forms as well. Because the disgust response is productive, composed of a variety of components that generate a cluster of inferences, expectations, and assumptions, it is possible for triggering entities to fit some components, but not others.[48] However, it is difficult to find triggering entities in which the *entire* cluster of expectations, assumptions, and inferences generated by disgust is satisfied, even in cases where the disgust execution system is not simply misfiring due to its hair trigger, but is performing one of its primary or auxiliary functions. This point can be elaborated by

---

[47] Compare this with cases of anthropomorphism discussed earlier, where talk of the projecting mind is useful and relevant when the folk psychological mechanisms are misfiring in reaction to clouds and animals, but rarely, if ever, is used to describe cases in which they are ascribing mental states and the like to people, who actually have them.

The case of race and racial cognition is equally informative. Amongst researchers concerned with race, there has been a consensus that races are not natural categories, that racial categories that group together people based on shared sensory properties like skin color do not pick out members who also thereby share a variety of other, deeper and more significant characteristics like socially, culturally, or morally relevant properties. Some have gone so far as to claim races simply do not exist (see Appiah 1995 for an eloquent defense of this position.) Such eliminativism flies in the face of our lived social experience, though, where races and racial distinctions seem to loom large. Claiming races do not exist thus creates an explanatory vacuum. If races do not really exist, why do we so easily see our social interactions in racial terms? Why, if it turns out there is no such thing as race, do we so persistently *think* there is? Psychological explanations will be a crucial ingredient in whatever complex story ends up filling this explanatory vacuum, and they will include appeals to features of the psychological mechanisms dedicated to racial cognition. See Kelly et al. forthcoming for a more detailed discussion.

[48] The same might be said of our folk psychological capacities and subtle forms of imperfect fit; anthropomorphism certainly comes in degrees. For instance, there is a difference between ascribing animacy and mentality to a cloud, and ascribing too much cognitive sophistication to a dog. In the later case, the dog probably satisfies some of the attributes ascribed to it, just not all of them.

focusing on properties of the psychology of disgust that were established in previous chapters.

Disgust is a Kludge

The Entanglement thesis has it that disgust is a kludge, created when a mechanism dedicated to monitoring food intake and protecting against poisons fused with a mechanism dedicated to monitoring for potential signs of disease and protecting against parasites. The disgust response is a piecemeal conglomeration of elements from each of these, and thus the response itself is not elegantly fitted to either poisons or parasites. The nausea produced in reaction to something infectious is superfluous, as is the contamination sensitivity produced in reaction to something that causes gastro-intestinal distress when ingested. Such superfluities, even in cases where disgust is performing one of its primary functions, create the explanatory role that is easily filled by saying that while one might be poisonous or another infectious, the full property of disgustingness is projected onto both of them.

On our revitalized understanding of projection, talking of projecting a property onto triggering entities is just a less precise way of rendering talk about the large set of assumptions and inferences that will implicitly be produced about how that entity will behave, affect the person, and should be treated. Once we see that the disgust response itself is an inelegant, piecemeal kludge, we can also see that the property of disgustingness understood this way, rarely, if ever, perfectly characterizes the entities that it is projected onto. Entities might be poisonous, and they might be infectious, but rarely will they be disgusting, and thus both.

That is, components of the productive disgust response have a psychological unity: they form a homeostatic cluster and are produced with nomological regularity, and activation of the disgust execution system reliably triggers the entire suite of components. These clustered components covary "in the head". However, we have no reason to think that the loosely corresponding properties "out there" in the world themselves form such a homeostatic cluster. In fact, starting with common sense and anecdotal report, and culminating with the Entanglement thesis, we have a variety of reasons to think that they do *not*. While an occasional entity may bear the complete set – all of the "poison" properties *and* all of the "parasite" properties – those properties by no means covary with any nomological regularity "out there" in the world, and so by no means form a homeostatic property cluster analogous to the corresponding psychological cluster of components "in the head". Thus, there will almost always be an imperfect fit between response and poisons, as well as an imperfect fit between response and parasites, the two best candidates for a good, seamless fit. And as a result of that imperfect fit, there will be explanatory work to be done by appeal to the projecting mind, in even these cases.

Disgust is Multifunctional

In addition to protecting against parasites and poisons, the Co-opt thesis holds that disgust was recruited to help regulate social interactions in a variety of ways, and has thus acquired a number of auxiliary functions as well. Nevertheless, as made clear in the discussion of byproduct hypotheses in chapter 4, when disgust is brought to bear on those auxiliary functions having to do with, for instance, social norms and monitoring ethnic boundaries, it brings to bear the full homeostatic cluster of components that make up the response. This creates more obvious forms of imperfect fit, as well as the cognitive

byproducts that are being explored in recent empirical research on moral judgment and disgust.

As that research has demonstrated, in such cases the mechanisms in the execution subsystem will project the property of disgustingness onto norm violators or members of vilified outgroups, who will be treated as if they were not just wrong or foreign, but tainted and contaminating as well. Here, most if not all of the clustered components of the disgust response will fail to be satisfied; the fit between response and triggering entity will be far from perfect. Nevertheless, because disgustingness is projected onto them, participants in the relevant social interactions will be treated as if they are contaminating, tainted, and dirty. Once again, due to the imperfect fit, there is explanatory work to be done by appeal to the projecting mind.

Disgust Allows for Significant Variation

Finally, the disgust acquisition subsystem allows for both individual and cultural level differences with respect to what triggers it. This type of variation opens up an explanatory role that appeal to the projecting mind can fill, but in a slightly different way than imperfect fit does.

One person might find meat delicious, while another person, who read Upton Sinclair's *The Jungle* at an impressionable age, finds it utterly disgusting. An easy way to talk about this is to say that the later person projects disgustingness onto meat, while the former does not. Seen this way, it does not seem to be an issue of who is right or wrong, but simply one of whose projective cognitive mechanisms are operating, and presenting the meat in a very vivid way, as nauseating, tainted, while the other one does not. The disgusted person does not find properties of offensiveness and contamination in

the meat and passively reflect them, while the steak eater fails to find and reflect them; the Upton Sinclair reader actively projects those properties onto the meat when they trigger his productive disgust mechanisms, even if, because those mechanisms operate implicitly and autonomously, he does not realize his mind is doing so. There are certainly nearby questions about whether each person is *justified* in his stance towards eating meat, but those are very different from the question about whether one person is correctly detecting a property in meat that the other is simply missing. Similar reasoning may be extended straightforwardly from idiosyncratic differences between individuals to patterns of cultural variation in disgust elicitors as well, and questions that can be raised about them.

In sum, then, it is not simply that its hair trigger generates a class of blatant false positives, nor merely the fact that the autonomous mechanisms underlying disgust are productive and implicit, that support the intuition that disgustingness is projected. It is, additionally, that there is almost always a role for psychological explanations to play when talking about disgustingness. This can be traced to other features of disgust – that it is a kludge, that it is multifunctional, and that its flexible acquisition system allows a high degree of variation – that create an explanatory vacuum. By virtue of generating an imperfect fit between response and elicitor in the case of the first two, and by allowing for substantial variation in the case of the last, it is nearly always natural to appeal to the projecting mind when describing cases of disgustingness.

5.6 The Master Argument Revisited

Let us return, briefly, to the Master argument against projectivism. As we saw, proponents of the Master argument were taking aim at positions fitting a generic form

composed of three main parts: a restrictive metaphysical view, a set of renegades made problematic by the imposition of that restrictive metaphysical view, and the appeal to projectivism to make sense of or account for those renegades. The Master argument claimed that the last move was illegitimate on the projectivist's own grounds, since projectivist resources are just as renegade as the properties they are being invoked to legitimate.

This argument gets no purchase on the revamped understanding of a projecting mind constructed here. First of all, at no point in time have we committed to anything resembling a restrictive metaphysical view. While our endeavor has been largely naturalistic in spirit, cleaving as close as it does to the cognitive sciences, it is not thereby held hostage to any such overarching puritanical position on ontology. While there are naturalistic philosophers of mind who hold such puritanical views (Fodor 1991), there are also naturalistic philosophers of mind with a much more liberal conception of what there is, and how questions of ontology relate to natural science (Laurence and Stich 1994, Stich 1996, Dennett 1991). Since there is no restrictive metaphysical view in play, no renegades are generated. The account given here anchors some properties such as disgustingness in the functioning of the mind, but it does not follow, nor have we concluded, that those properties do not exist, that nothing is really disgusting or that all statements ascribing disgustingness are false. (Indeed, I have said *nothing* about the semantics of such claims). More generally, our view is not tacitly committed to, nor have we explicitly endorsed, any more encompassing position that makes the mind particularly problematic or unintelligible.

Second, we have shown how talk of a projecting mind can be construed as a convenient shorthand for the kinds of theoretical talk employed by cognitive scientists when they are characterizing fairly autonomous components of the mind, specifically the operation of implicit and productive mechanisms that are easily activated. This account is both interesting and clearly coherent on the face of it. Since there is no prima facie reason to think otherwise, and since the Master argument gets no traction on it, we may conclude that this form of projectivism is itself coherent.

Certainly there are broader problems having to do with integrating mental properties into the world order and elucidating their relationship to physical properties. To claim otherwise would amount to dismissing the core issues in the philosophy of mind. In claiming to have avoided the Master argument against projectivism, we do not commit this absurdity, and neither do we claim to have thereby solved every facet of the mind-body problem. We have simply shown that the idea of projectivism can be made sense of. Moreover, in anchoring the notion in the functioning of autonomous, implicit and productive mechanisms, the common currency of explanations in cognitive science, we have put the notion of projectivism on the same footing as one of our most vibrant and flourishing natural sciences. Therefore, there are no immediate grounds for singling out the projectivist-cum-cognitive scientific resources developed above, or for condemning them as uniquely problematic, unaccountable, or unintelligible.

5.6.1 Objections and Replies

We can quickly respond to a few of the worries that might come to mind upon first encountering the above account.

Objection 1: Your conception of projection is invalid because it misrepresents Hume (or some other historical figure).

Reply: Had my main aim been Hume exegesis, I would agree that my efforts have fallen woefully wide of the mark. Luckily that has not been my main aim. Rather, I have been primarily interested in extending the tradition in which Hume worked, not correctly explicating the nuances of his position.

Interestingly, however, I am not so sure that Hume would be terribly unhappy with the view sketched above. He has been proven remarkably prescient when it comes to anticipating the broad themes of cognitive scientific research. Consider his discussion of anthropomorphism, where he eloquently describes the phenomenon, even if he lacks a detailed understanding of the secret springs that generate it:

> "There is a universal tendency among mankind to conceive all beings like themselves, and to transfer to every object, those qualities, with which they are familiarly acquainted, and of which they are intimately conscious. We find faces in the moon, armies in the clouds; and, by a natural propensity, if not corrected by experience and reflection, ascribe malice or good-will to everything, that hurts or pleases us…trees, mountains and streams are personified, and the inanimate parts of nature acquire sentiment and passion"
>
> (Hume 1757: 29, as quoted in Atran 2002, page 68).

Objection 2: This is not how projectivism is understood or used in the current literature.

Reply: Fair enough, but my account amounts to a new way of unpacking and understanding the core slogans, metaphors, and catchphrases associated with projectivism. As I was at pains to show in Section 2, those metaphors have been flushed out in a variety of ways over the course of time, all interesting, but none definitive. I maintain that while my account is certainly novel in many respects, it nevertheless remains true to many of the ideas at the heart of the tradition. That my account differs from other, perhaps more linguistically oriented notions of projection, is an indication

that I have indeed made progress, or at least done what I set out to do – produce an account that is original and distinctly grounded in psychology.

Objection 3: Your account cannot do all of the work that projectivism is supposed to or has been called upon to do. Therefore, it is not worthy of the name.

Reply: While I disagree with the conclusion of this objection for reasons just stated, I suspect the premise is correct, and that the account given cannot do all of the work that projectivism has been called upon to do in the past. The details of how much philosophic work it can do, on what fronts it succeeds and where it falls short, and so forth, remain to be seen. It is altogether possible that old versions were better fit to solve different problems, or that other philosophers waving the projectivist banner were overreaching in their pursuit of their own ambitions. My account and aspirations may turn out to be considerably less grand than those previously put forth, but at this point, I take that neither to be a point in my favor, nor against me. For now, one insight my work might be taken to support is that the notion of projection is not necessarily confused or incoherent, and certainly deserves a place at the table in serious philosophic discussions. My account also suggests that while it is not confused, it may not be exactly what it was once thought to be, either.

Objection 4: You have not shown the limits of your view by indicating how many other sorts of cognitive mechanisms fit the above description enough to say that they project properties.

Reply: This is true, but it is also not something that can be determined from the armchair. As cognitive science progresses, it may uncover and delineate a large variety of autonomous, implicitly operating, productive mechanisms. Further research might reveal

them to be the sort of multifunctional kludges that admit of individual and cultural level variation, as the disgust system turned out to be. If so, then there will be explanatory work for the idea that the associated properties are projected onto the world by an active mind. At this point, it is hard to say what the empirical work will reveal, though, as mentioned above, there have been some promising indications of late (Wilson 2002, see also Marcus forthcoming). If I were forced to place a bet right now, I would put money on these types of mechanisms being far from rare.

Objection 5: There is an elephant in the room that you have not addressed. Do you think morality is projected? Can your account capture moral properties like badness or rightness?

Reply: These are very big, very interesting, very difficult questions, and obviously they need to be handled with great care. For now, it will have to suffice say that there is no obvious reason to reject affirmative answers out of hand.

Much of the recent empirical work on our moral psychology is uncovering a wide array of implicit and autonomous cognitive mechanisms, and many of those appear to be productive (see Haidt 2001; Nichols 2004; Hauser 2006; also see Doris and Stich 2005 and Nado et al. forthcoming for overview and discussion). Additionally, there has been suggestion from more than one quarter that many of those mechanisms are kludges, or initially evolved for some other purpose before being co-opted to the roles they now fill in producing moral judgment and moral motivation (Nichols 2001, Stich 2006, Knobe forthcoming). As in the case of disgust, it appears they often perform their new functions imperfectly. A psychologized projectivism may be a quite attractive way to account for

the corresponding properties that such mechanisms purport to detect "out there" in the world.

Since we are speculating, though, it should be emphasized that on the assumption that the projectivist account I've sketched can be extended to moral properties more generally, it will not be the end of the world. No form of moral nihilism or eliminativism would follow, nor would any conclusions about unintelligibility or massive error in moral discourse. My form of projectivism anchors properties in the functioning of the mind, but it does not entail that such properties do not exist, that nothing in the world really bears them or that all statements ascribing them are false.

Objection 6: What about qualia? Projectivist theses are often associated with qualia such as colors; your account does not seem to be about such properties at all.

Reply: The issues surrounding colors, and qualia in general, are complex and difficult, and the account offered above certainly has no straightforward application to those debates (though we will briefly take up the color analogy in the next chapter). Color experience appears to be utterly and purely subjective. Unlike the sorts of cases involving disgustingness or anthropomorphism we focused on above, color experience does not involve a cluster of attendant inferences, assumption, expectations, emotions, intentions, dispositions to action in any particular way, or anything else we might get at in an experimental setting. With color experience, we are still left wondering: what exactly might it mean to say we are projecting, say, blueness, and what is it, exactly, that we are projecting? In cases of color, therefore, the Master argument still has some bite.

It is worth noting that even in speculating about the relevant cognitive mechanisms and the phenomenal character of experiences that often accompany their

operation, nothing I have said above assumes anything more than a *correlation* between mechanism and phenomenal experiences. The existence of such systematic correlations is common ground for most staked out positions on qualia and consciousness (substance dualist, epiphenomenalist, emergentist, property dualist, new wave materialist, etc.); the positions differ on how best to account for it (see Chalmers 2003, McLaughlin forthcoming).

Some may still take the silence of our account as a failing of the treatment of projectivism. From a different point of view, though, it can be taken as a virtue that we have made sense of the idea of a projecting mind apart from the debates about color, and without becoming mired in the quagmire of qualia. Indeed, we can now give a concise statement of what projectivism amounts to and when such explanations are appropriate that is not deeply enmeshed with the mysteries of consciousness. We can instead make sense of the notion in the context of broader, perhaps more tractable problems about how the rest of the mind works. Part of what I set out to do is secure a way of talking about projectivism which does get bogged down in or held hostage to the debates about consciousness and qualia; I do claim to have made some strides towards that goal.

Objection 7: You never gave a definite answer to one of the most intuitive questions one might ask here: is anything *really* disgusting?

Reply: Intuitive questions are not always good ones; in answer to this one: yes and no.

5.7 Conclusion

Much recent debate about projectivism has been couched in terms of semantics and truth conditions, or concepts and content. But the terminology and imagery of projectivism has a history that stretches back farther than the modern linguistic turn.

Indeed, the slogans and images at the heart of the tradition become much more appealing, and to some extent, appropriate, once we start thinking in terms of the general explanatory format used in cognitive science, and of the mind in terms of a collect of fairly autonomous, productive, and implicitly operating cognitive mechanisms. Using the tools and vocabulary of cognitive science, we have updated an old philosophic idea about the mind, and shown how it is useful in accounting for phenomena like anthropomorphism, and can be used to shed light on the intuition that properties like disgustingness are projected onto the world. With this new, psychologized understanding of projectivism in hand, we have shown that it avoids the argument advanced by many philosophers who are hostile to the tradition, and answered some of the knee jerk objections that it might provoke.

In the next chapter we look at other accounts of disgustingness that have been proposed, and point out where they fall short. We can end this one by summarizing some of the more interesting conclusions also fall out our new conception of projectivism:

1. Whether or not a particular property is projected is a question about what the mind is doing when it presents that property, and is therefore an empirical question.

2. As we saw from the case of anthropomorphism, mentalizing, and the folk psychological capacities involved, no conclusions about the status of the triggering entities immediately follow from the fact that the mind is projecting some property.

3. As we saw in the case of disgust and disgustingness, questions about the status of the projected properties can be greatly informed by appeal to both proximate and ultimate explanations of the components of the mind that are involved, whether they are "designed" to accommodate variation, whether they are multifunctional, and whether they are elegant machines or kludges, and how such facts effect the degree of fit they have with their triggering entities.

4. The degree of fit between a response and its object affects the pragmatics of explanation, making projectivist explanations more natural and appropriate in some cases than in others.

Chapter 6: *That's* Not Disgusting! A Critique of Three Views of Disgustingness

6.1 Introduction

In this final chapter, we will examine three different accounts of the property of disgustingness. The stance taken will be mainly critical, and so the positive, projectivist account developed in the last chapter will be present only tacitly. In Section 2, we will quickly motivate the types of questions that the three accounts serve to answer, and say something about where such accounts sit in relation to the larger philosophic landscape. Section 3 distinguishes two closely related accounts of disgustingness, each inspired by functionalism and functional accounts of mental properties in general. Separating out the subtle differences between those two accounts requires that we take a step back and return to the foundations of functionalism. After criticizing those functionalist views, Section 4 turns to fittingness accounts of disgustingness, focusing on recent work in metaethics on sentimentalism. The section ends with a criticism of fittingness accounts that again draws on the features of disgust established in earlier chapters. We end with some concluding remarks that highlight the major themes running through the individual criticisms.

6.2 Location Problems and their Solutions: Democritianism, Profligate Realism, and Points In Between

Is anything *really* disgusting? When two people, or two cultures, disagree about whether something is disgusting, is one of them right and the other wrong? Can something be said to be disgusting independently our capacity to be disgusted by it?

What is the nature of the link between the emotion of disgust and the disgusting things that elicit that emotion? What is the nature of the property of disgustingness itself?

One might not think these questions are all that interesting in themselves, and I might not disagree too vehemently. But they become a little more intriguing when placed in a larger context, and viewed as specific instances of larger questions having to do with reality and realism. Consider a very Crude Democritian picture of what exists: all there really is in the universe are atoms and the void, and perhaps the laws that govern the motion and interactions of the atoms (or whatever fundamental particles physics finally settles upon). On this view, all things are just atoms arranged in different ways; there aren't any ghosts, but there aren't really any chairs, marriages or nations, either. Mountains and minds don't actually exist, there really is no such thing as beauty or value, and certainly nothing is really disgusting. There are only atoms and the void.

Think of the Crude Democritian picture as one pole along a spectrum. At the other extreme would be an equally crude picture we might call Profligate Realism. On this caricature of a view, chairs, marriages, and nations are just as real as atoms, and minds exist in the same way as mountains. In addition to those things, there really are properties like value and beauty, and their existence is as ontologically independent of the rest of creation as everything else. Each is a fully objective property, completely independent of human beings and their psychological apparatus, and can be specified without reference to humans, their institutions, social practices, or individual responses. The property of disgustingness is likewise an objective, independent component of reality. Some things straightforwardly bear that property, and others straightforwardly do not. And so some things really are disgusting, and some are not. Different people might

be better or worse at detecting that property, but that is merely epistemic; when two parties disagree about whether something is disgusting, one is right and one is wrong, end of story.

Needless to say, few philosophers who worry about these sorts of ontological questions would be happy with either Crude Democritianism or Profligate Realism. It is the rejection of either of them and the simple answers they provide, however, that make the questions we started with all the more pressing. In and of themselves, those questions specifically about disgustingness might seem merely academic, in the pejorative sense of the term. With the proper context, we can see that by giving an account of disgustingness, one might thereby develop a way of accounting for others – perhaps even those that make life worth living, like beauty and value.

The difficulties in accounting for any of these can be gathered together under the heading of "location problems". These make up a core area of contemporary metaphysics, which investigates the relations between certain properties, such as colors, moral and aesthetic properties, on the one hand, and (usually) the entities and processes posited by the natural sciences on the other. The problem is one of locating those properties in the natural world, of being able to say something about how they are related to the picture of the world given to us by natural science. The overriding concern is to figure out where we might locate these entities and properties in relationship to the rest of nature.

Such questions have been asked about the ontological status of many different entities, including colors, intentional properties, qualia, social facts, numbers, fictional objects, rights, and so forth. Projects concerned with illuminating the relation of such

entities to better understood areas of nature go by a variety of names. Some philosophers tend to think of themselves as examining the prospects for naturalizing the set of properties or entities in question. For instance, Fodor worries that "there is no place for intentional categories in a physicalistic view of the world; that the intentional can't be naturalized (Fodor 1987, page 98). Others favor the terminology of location: "[T]he central problem is that locating mind with respect to the physical world (Chalmers 2003, page 1). Simon Blackburn agrees with Frank Jackson in claiming "[W]here there is something that threatens to transcend the physical or the natural, the way to demystify it is to "locate" it in the natural order"; moreover, he suspects he speaks for the majority: "[M]any writers would agree with Jackson that a fundamental task of metaphysics is what he calls the location problem: showing how to locate the mystifying area in the natural world" (Blackburn 2000, page 119-20).[49]

The trouble with the Crude Democritianism and Profligate Realism that we started with is that neither of admits of any shades of gray. As one might suspect, philosophers have spilt much ink in attempts to chart out the vast middle ground between those two extreme poles, formulating intermediate positions and constructing ways to account for where things like nations, values, and beauty fit into the greater scheme of things.

---

[49] While philosophers of many stripes seem to agree that location problems are central to many contemporary philosophic debates, attempts to formulate such problems with any degree of precision are difficult to find. Indeed, there is a cluster of related meta-philosophical questions which are rarely commented upon, but on which there appears to be much less agreement. These questions concern issues about the correct methods to use in pursuing solutions to location problems, and the appropriate standards for success in solving them. Debates about the role of intuition, the viability of conceptual analysis, and the relationship between findings from both cognitive science and other natural sciences, on the one hand, and philosophy, on the other, can all be seen as circling such issues, if not addressing them head on.

A complete survey of all possible intermediate positions would require a not small book. Luckily, our focus on the property of disgustingness will help confine the discussion to three distinguishable accounts that have been given of it. I do not think any of them succeed, and after sketching each account, I will try, concisely to say why. Once again, in doing so, we will take for granted that the cognitive and evolutionary underpinnings of the emotion of disgust defended in the first half of the dissertation is correct.

Before we dig in, a word on terminology will be useful. Many simply assume that it goes without saying that a property like disgustingness is dependent on disgust and the disgust response in some way, and unless you are a Profligate Realist, it is hard to disagree. The term "response dependent" is a fashionable buzzword right now, and probably because of this, it is far from univocal. If anything, it denotes a large family of views, many of which have important and interesting differences.[50] Indeed, all three of the views under consideration below take the property of disgustingness to depend on the disgust response, in one way or another. They differ largely in how they construe the nature of that dependency relation, or what conclusions to draw from the construal. Therefore, I will attempt to steer clear of that terminology in what follows. Where it is unavoidable, I mean it in whatever way it is used by the authors I am currently discussing, and that will be made clear in the local context.

6.3 Functionalism and the Color Analogy

---

[50] Merely claiming that the identity of a property depends in some way on our response to it does not resolve location problems, either. Consider a paradigm example of a candidate response dependent property, color. A cursory glance at the literature on color reveals many lively debates. For instance, Byrne and Hilbert (2003) distinguish 5 major families of views about the proper way to understand the place of colors in nature. All of these views appear to agree that colors are response dependent, or that colors can only be identified by reference to the associated responses of our visual apparatus. Each position draws different conclusions from this fact about the location of color, however.

The first two accounts of disgustingness we will consider are broadly functionalist in nature. They differ in subtle respects, and in order to clearly understand how, we will return the roots of functionalism to see where the two different versions diverged as they were developed.

Versions of each type have been advanced in a nearby area of philosophic inquiry, namely the philosophy of color. The analogy to color is useful for a couple of reasons. First and most generally, since other authors have already developed rather detailed functionalist accounts of color, we may borrow, where appropriate, the structure and sophistication they have achieved, rather than having to begin from scratch. Another reason is closely related to this. Unlike the beliefs, desires, and other mental states that functionalism was initially developed to account for, colors do not seem to be properties that are "in the head". Rather, correctly or incorrectly, they appear to be properties of the things "out there" in the world that are the objects of our perception. Thus, the analogy to color provides a nice model of how to apply functionist lines of thought to another property that, correctly or incorrectly, appears to inhere in objects "out there" in the world, namely disgustingness.

A final reason the analogy with color is useful is the role the natural sciences now play in the philosophic debates. A noteworthy feature of the philosophy of color is that it is no longer possible to fruitfully engage in the debates over color realism and the location of color in nature without knowing a good deal of the relevant science of vision. This trend can be traced back to the work of C. L. Hardin and his groundbreaking book *Color for Philosophers* (1988). In the recent words of another philosopher of color, Jonathan Cohen:

> "Prior to the publication of [Hardin 1988], philosophical work on color had been conducted in roughly the same terms in which it had been carried out by the famous moderns – Galileo, Boyle, Locke, et al. But once Hardin pointed out that a vast field of empirical research had developed since the modern period, and showed convincingly that these developments impose serious constraints on ontological and epistemological disputes about color, the philosophic landscape was forever changed. Subsequently, philosophic work on color has increased dramatically in both sophistication and interest"
>
> (Cohen, draft, page 1)

It will come as no surprise that I view this infusion of empirical data into the philosophic debates as a welcome mark of progress. It is doubly congenial in that it too provides a model of how to pursue our interest in disgustingness. For, we now know much more about the science of disgust, about the mechanisms and processes that underlie the disgust response. Just as the philosophic debates over color were enriched by an infusion of the relevant science of vision, the philosophic debates over a different sort of property, disgustingness, can now be enriched by an infusion of the relevant empirical work as well. But first, a look at functionalism from a broad perspective will help set the stage for digging into the specifics of the two accounts of disgustingness.

6.3.1 A Brief History of Functionalism

The mind/body problem was initially formulated by Descartes, who went on to offer a solution now called substance dualist: mind and body are made of distinct substances, each of which can have causal influence upon the other. Substance dualism has few adherents today. Instead, materialist positions like behaviorism, identity theory, and functionalism all attempt to answer the same question about the relationship of mind and body without appeal to an entirely distinct ontological domain of the mental. These positions seek to understand the relationship of mental states to physical states, but are concerned to do so in materialist terms.

According to behaviorism, the mind is really behavior. Behaviorism construed talk of mental states as abbreviated talk about publicly observable behavior, patterns of behavior and dispositions to behave, and held that descriptions couched in mental terms are ultimately translatable to descriptions about behavior and the movement of the body. Thus, mental states are mere aspects of behavior, and the mind is not genuinely separate from the body, but an aspect of what it does (see Ryle 1949).

The main objection to behaviorism is fairly obvious. In failing to countenance anything but publicly observable behavior, it clashes rather flagrantly with our sense of having an inner life, and with the robust intuition that the mind is an inner cause of behavior rather than an aspect of it. Indeed, mind and behavior seem easily separable and quite distinct from each other (see Putnam 1968, Dennett 1978).

Alternatively, identity theory holds that the mind is really the brain. The hope of identity theory is that mental states can be identified with certain physical states, specifically with brain states. Just as modern chemistry discovered that water is identical to H20, so too will modern neuroscience reveal that mental states, more specifically mental state *types*, like pain, the desire for a beer, or the belief that it will rain tomorrow, can be identified with brain state *types*, like the firing of C-fibres (the usual philosophic placeholder for some predicate in a mature neuroscientific theory). While the identity of the two will not be demonstrable a priori, nor could one type of concept be derived from the other, empirical research will show how the two types of categories systematically map onto one another. In virtue of this identity, mind and brain will be shown to be the same thing (see Place 1956, Smart 1959).

The main objection to identity theory is that it is overly chauvinistic. In binding the identity of mental states so tightly to the actual brain states of human beings, identity theory makes it impossible in principle for animals with brains significantly different from our own to have mental states at all. It is highly counterintuitive that animals like squids or bats are not the subjects of mental states such as pain, or cannot think at all. Moreover, identity theory makes it equally impossible for computers or creatures with significantly different chemical composition that ourselves to have minds, for much the same reason. If mental state type X is identical with the human neurological state type Y as discovered and specified by mature neuroscience, and some other creature does not have any neurological states of type Y, then it immediately follows that the creature does not have any mental state of type X (see Putnam 1973).

Functionalism was the positive view that grew out of the dissatisfactions with both behaviorism and identity theory, but the most intuitive way to state the core idea is in terms of the computer metaphor: minds are to brains as computer programs are to the computers that run them, as software is to the hardware it runs on. According to functionalism, mental concepts are functional concepts, and thus mental states are functional states. The concepts specify a state in terms of whatever role it is playing within a larger cognitive system, but remain silent about the physical processes that are performing the functions. Mental states are identified by what they *do*, rather than the specific type of physical stuff that is doing it.

Functionalism has ready responses to the main objections leveled against both behaviorism and identity theory. Unlike behaviorism, it does not identify mental states only with patterns of behavior and dispositions to behave. Rather, in specifying mental

states in terms of their role in an entire cognitive system, it identifies them by reference to the functional relations they bear to behaviors, to perceptions, to other mental states, and on some views, even to objects in the environment. Functionalism is thus able to begin distinguishing the mental states from the behavior they cause, and countenance the intuitions about the rich inner life of the mind (though many philosophers claim that fully accommodating qualia and the first person character of the mental is an insurmountable stumbling block for functionalism (see Block 1980, Nagel 1974, though see Dennett 1991 for an opposing view)).

Unlike identity theory, functionalism does not bind mental state types to any specific physical state types, or even to any particular type of physical stuff in general, be it organic, metallic, silicon based, or anything else. Just as a single type of program can be run on a variety of different computers, so too can mental state types be run on a variety of different "hardware": human brains, animal brains, Martian brains, groups of people, and even, in principle, powerful computers. While each individual *token* of a mental state is identical to or realized by some *token* physical state or other, taxonomies of mental states and taxonomies of physical states will not map onto each other in any systematic way. Instead, mental state types will cross cut physical state types, and vice versa. Intuitively speaking, functional states, and thus, according to functionalism, mental states, are multiply realizable.

The strategy for avoiding the common objections to behaviorism and identity theory, the inspiration drawn from the computer metaphor, the specification of mental states in terms of their role in the larger cognitive economy, and the association of mental state types with those functional roles – these are the broad themes shared by

functionalist views. Upon closer examination, however, those views bifurcate into two importantly different families, namely filler functionalism and role functionalism. The differences between these two will be developed in detail as we look at filler and role functionalist accounts of disgustingness, but an analogy will help set the stage.

Consider the difference between the role of Hamlet and an actor, say Kenneth Branagh, who plays that role in a staging of the play at the Globe Theatre in London. Clearly the two are not the same thing. There is a distinction between a role and the thing that fills it. Kenneth Branagh is a person, an actor who has played a number of other acting roles, but who has many other characteristics besides, some perhaps essential to who he is, others not. He is flesh and bone, the son of a specific mother and father, has a particular biochemical makeup, and a set of memories of his childhood, adolescence, and the rest of his unique personal history. He also has red hair, and was once married to Emma Thompson, with whom he made an enjoyable movie version of *Much Ado About Nothing*.

The role of Hamlet, on the other hand, is an abstraction. It can be filled at different times, in different places, and by different people, but it can only be filled when Shakespeare's play *The Tragedy of Hamlet, Prince of Denmark* is staged. That thing, the abstract role of Hamlet, is defined by reference to the lines the character speaks and the relations the character bears to the action, events, and other characters within the larger play: he is the brilliant but angst ridden Prince of Denmark, the indecisive friend of Horatio, conflicted son of Gertrude, accidental murder Polonius, and lover of Ophelia. In playing Hamlet, individual actors like Branagh can have, at specific places and over short periods of time, an intimate relationship to this role, namely playing it, occupying it, or

filling. And of course, some actors fill it better than others. But not even the best of them thereby becomes identical to the role itself; none of them *is* Hamlet.

Returning to functionalism, it is their opposing stances towards the significance of this type of distinction, between a role and the thing that fills it, that separates filler from role functionalists. Both attempt to do justice to the insight that mental states are closely associated with functional roles, but they do so in different ways. Simply put, filler functionalists hold that mental state types are identical to whatever physical type fills the relevant functional roles, and role functionalists are hold that mental state types are identical to the abstract functional roles themselves. In terms of our analogy, filler functionalists think mental states are more like Kenneth Branagh; role functionalists think they are more like Hamlet.

Delving into the minutia of the differences between these views would take us too far a field, and the most relevant differences will be flushed out below anyway. Nevertheless, Table 1 presented below systematizes some of the main differences between filler and role functionalism, between how they were initially motivated, between the answers they give to critical questions, and the main weakness of each. The reader is invited to refer back to it, where needed, as we proceed with our discussion of disgustingness.

Differences Between Filler & Role Functionalism

	Filler Functionalism	**Role Functionalism**
Common Aliases	Realizer or Occupier Functionalism; Analytic Functionalism; Categorical Basis Views;	Machine Functionalism; Psychophysical Functionalism; Dispositionalist Views;
Spiritual Affinity with	Identity Theory	Behaviorism
Variety of Materialism	Reductive	Non-Reductive
Pioneering Philosophers	David Lewis (1966, 1972), David Armstrong (1968)	Hilary Putnam (1967), Jerry Fodor (1968, 1997)
Initial Versions Picked Out Relevant Functional Roles via	Conceptual analysis and the platitudes of our folk psychology, how they specify connections between common sense mental states and their role in guiding behavior	Empirical research in cognitive science, which identifies the "program" and associated machine states of the human mind through careful experimentation
Construed Mental Predicates as Identified Mental State Types with	Non-rigidly designating 1st order physical properties Whatever physical properties *occupy* or *fill* the relevant functional role	Rigidly designating 2nd order functional properties The functional role *itself*, i.e. the abstract set of relations to other states and behaviors that constitute the role
Construed Mental State Types as	First order properties, i.e. those properties that have the relevant second order property of filling a functional role	Second order, relational properties, i.e. the property of having some property that meets some functional specification
Multiple Realizability	Limited: results in local, species specific type-identities	Robust: limited only by what types of physical properties can, in fact, fill the relevant roles[51]
Common Objections	Difficulty accommodating causal generalities	Difficulty with mental causation, accommodating intuitions about the causal efficacy of the mental

Figure 6.1

This table charts out the subtle differences between filler and role functionalism, and the characteristic answers each gives to some specific questions about the relationship between the mental and the physical. The information was drawn from various papers, which can be consulted for more detail, including Jackson (1996), Bennett (2007), McLaughlin (2003, forthcoming), Goldman (2007), as well as papers and overviews found in Chalmers (2002).

[51] For instance, a chimpanzee cannot fill the role of Hamlet. This is a matter of fact but not principle.

6.3.2 Filler Functionalism: A Categorical Basis for Disgustingness?

I am not aware of any philosopher who has developed a filler functionalist view explicitly about disgustingness in any detail (though Goldman (2007) briefly considers what one would look like before he rejects it). Nevertheless, we may easily construct one by using the color analogy, and building on filler functionalist views of color. In particular, the one we are building presently is closely modeled on Brian McLaughlin's view of color (2003). The basic idea of this filler functional account is that the property of disgustingness is to be identified with whatever physical property plays *the disgustingness role*, if there is such a physical property. If such a property can be identified, then entities, actions and the like that bear that physical property are disgusting.

Similar to the familiar distinction between colors and the phenomenal or qualitative character of our color experiences, so too can a distinction be made between disgustingness and the phenomenal or qualitative character of our experiences of disgust. Indeed, disgust is an emotion, and, as opposed to experiencing the presence of a color, the experience of having an emotion is comprised not just of its phenomenal components, but of various other cognitive, affective, and behavioral components as well. Disgustingness is putative a property of entities, actions, and the like, that induce this emotion. This property is to be distinguished from not just the cognitive, affective, and behavioral components of the response, but from the phenomenal component of the emotion as well, i.e. what it's like to be disgusted by such entities, actions and the like.

The average man on the street is quite probably ignorant of the nature of disgustingness, but he forms a conception of disgustingness by its relation to his

experience of being disgusted. In other words, being disgusted by something involves an unknown cause, the property of disgustingness, and a suite of effects, one of which is the known experience of being disgusted. We can formulate this idea as follows:

> Basic Filler Functionalist View: disgustingness is that property which disposes its bearers to induce disgust in ideal agents in normal conditions, and which must (as a matter of nomological necessity) be had by everything so disposed.

The view, like analogous views about colors, includes a functional, thus topic neutral, analysis of the phenomenology and concept of disgustingness. That analysis is taken to fix the referent of that concept, but also express a condition that is necessary and sufficient on satisfying it.

According to the filler functionalist view, disgustingness is a functional property in that it plays a particular role in relation to our emotions and other responses, cognitive, affective, behavioral and otherwise. The functional role it plays is that of being the property that disposes its bearers to induce the emotion of disgust in standard agents in standard conditions. Moreover, it is the property that (nomologically) must be had by everything so disposed. As shorthand, we can call this "the disgustingness role". Putting the pieces together, since any property that fills the disgustingness role satisfies a condition that is necessary and sufficient for being disgustingness, then any property that uniquely fills the disgustingness role *just is* disgustingness. Likewise, being disgusting just consists in having a property that fills the disgustingness role.

A few features of this view bear comment. First, in construing disgustingness as a property born by entities, actions and the like, the proposal is broadly consistent with the phenomenology and experience of disgust. That experience presents disgustingness as a property of the entities, actions and the like which induce disgust, i.e. as a property that

we detect in the world around us and react to accordingly, rather than, say, as a property of our reaction.

Second, as the name makes clear, this view would identify the property of disgustingness with whatever physical property fills the disgustingness role, rather than the more abstract disgustingness role itself. In doing so, it depicts disgustingness not as a disposition, but as a *categorical basis* of a disposition. It thus imposes fairly stringent requirements on candidate physical properties. If any physical property is to be identified as the property of disgustingness, it must be common to *all* fillers of the disgustingness role, and it must also be the *unique* physical property that all fillers have in common. If either of these requirements is not satisfied, then it will turn out that nothing is, in fact, disgusting.

Third, the appeal to standard agents and standard conditions immediately leaps out as problematic, especially in light of the well-known individual and cultural variation of disgust. This will be addressed in more detail below.

Finally, one particularly interesting feature of this proposal is the role it allows for cognitive science in determining what, if anything, fills the disgustingness role. In this, the functionalist proposal can be seen as a clear exemplar of the project of specifying the essence of the property of disgustingness if there is one.

6.3.2.1 Objections to a Filler Functionalist Account

Two main difficulties can be raised to unvarnished filler functionalist accounts of disgustingness. Both arise from places where the analogy between color and disgustingness breaks down. The first is relatively straightforward, and stems from the appeal to ideal agents and normal conditions. Unlike with color, there is substantial

variation in the elicitors of disgust, both between individuals and between cultures. Indeed, many of the auxiliary functions that disgust plays take advantage of this feature of the system for strategic purposes. In allowing disgust to be directed at members of other cultural groups, or at transgressors of certain local norms, it helps sustain certain social dynamics and cultural patterns that arise for some of the strategic purposes discussed in Chapter 4. Thus, the appeal to ideal agents and normal conditions begs the question as to which individual or cultures are ideal, which conditions are normal. Nor do there seem to be any grounds for deciding which individual or cultures are normal or ideal that would not manifestly chauvinistic.

The character of the mechanisms underlying disgust gives further support to this objection. Unlike the systems underlying color vision, the disgust system is equipped with a flexible, open-ended acquisition system. Mechanisms in this system allow agents to *learn* what is disgusting and what is not, and what is learned often differs from one culture to the next. While there is innate structure in the disgust system, it also includes a variety of mechanisms devoted to the acquisition of elicitors, and this gives the system a flexibility that is not present in the color system.

Most problematic, however, is the requirement that if any physical property is to be identified as the property of disgustingness, it must be common to *all* fillers of the disgustingness role. One of the most striking things about disgust is that there is enormous diversity in the sort of things that elicit it, from rotting meat to pus-oozing sores to violations of certain social norms to the ethnic markings of rival tribes. This gives strong reason to suspect that there is no property shared by all elicits of disgust, other than that they trigger the emotion, and thus that nothing is disgusting (at least on

this view of disgustingness).[52] This suspicion could be wrong, of course; common sense is not always the best guide in these matters. Entities and substances with greatly different surface appearances, such as coal and diamond, or glass and sand, have turned out to have common physical bases. But in the case of disgustingness, we have additional reason to think there is no physical basis shared by all elicitors. For, the combination of the Entanglement thesis and the Co-opt thesis cast further and more principled doubt on the idea that this emotion is doing anything like keeping track of a single type of physical property through a variety of different manifestations. Rather, disgust is performing a variety of functions, and in so doing it is attending to a variety of different properties.

Filler functionalism would seem to lead to eliminativism about the property of disgustingness. In and of itself, this is not a terrifically disturbing conclusion, but it is not altogether satisfying, either. Without further explanation, the claim that all ascriptions of disgustingness are false sounds absurd; anyone who has had to change a nasty diaper or been near a garbage dump on a hot and humid summer day would probably beg to differ. Perhaps a defender of a filler functionalist account of disgustingness could find a way to avoid the conclusion, or make it more palatable. For right now, though, there does not appear to be any such defenders, so we cannot evaluate such maneuvers. Instead, we will turn our attention to other accounts, which might be more congenial.

6.3.3 Role Functionalism: A Disposition to Elicit Disgust?

[52] Goldman (2007) objects to a filler functionalist account of disgustingness on similar grounds: "If disgustingness is the ground of the disposition to elicit feelings of disgust, mustn't it be a property common to all disgusting things…? Is there any such common ground? Cognitive science findings make such a thesis problematic" (Goldman 2007, page 6).

Role functionalism about disgustingness enjoys the explicit support of at least one philosopher, namely Alvin Goldman. His discussion of the view is done in the service of larger, methodological points about naturalizing metaphysics that he is more directly occupied with. In that discussion, however, he is motivated by some of the concerns just mentioned: while it seems unlikely that there is a common ground to all fillers of the disgustingness role, the eliminativist conclusion that follows from a categorical basis view is far from satisfying. Instead of embracing such a conclusion or tinkering with the view that leads to it, he recommends instead a dispositionalist or role functionalist account, that better steers the course between robust (or naïve) realism and eliminativism: "So let us consider the alternative approach: dispositionalism or response-dependence. On this view, disgustingness is simply the disposition to produce a disgust response in humans" (Goldman 2007, page 6).

This view identifies the property of disgustingness with the disposition to induce disgust, rather than the absolute basis of that disposition. Disgustingness is the role itself, a more abstract set of relations to our mental states, affective responses and behaviors, rather than any physical property or properties that fill it. Disgusting things are just those things that are disposed to elicit the disgust response.

In identifying disgustingness with the role rather than what fills it, dispositionalism avoids both of the stringent requirements that made a filler functionalist account unappealing. Dispositionalism holds that since disgustingness is the role rather than the occupier of the role, anything or any physical property that bears the second order property of occupying the disgustingness role is thereby disgusting. Disgustingness is the role itself, and different things and properties are disgusting simply in virtue of

occupying it, of having that second order property; no unique common ground is required.

Role functionalism turns the focus of the account from the categorical basis of dispositions to the dispositions themselves. In this case, the disposition is specified by reference to the disgust response. Goldman turns his attention to that response, to further refine his view. Though he does not use our terminology, he is driven by considerations captured by the Entanglement thesis to suggest, in a revisionary spirit, that we break the property in two: one that is the disposition to trigger the poison mechanism (the affect-program responses), and one that is the disposition to trigger the parasite mechanism (the core-disgust responses). Thus, we end up with a view that countenances some of the subtleties of the disgust response, avoids the problem of common ground, and yields the properties of disgustingness$_1$ and disgustingness$_2$.

> "Returning now to the dispositional approach to disgustingness, exactly which responses are constitutive of this disposition? Are they <u>all</u> constitutive of it? Here is a different way to approach the problem. Since we are already dealing with revisionary accounts of disgustingness, why assume there is a single property? Why not distinguish different disgustingness properties, so as to cut nature better at its joints? One way to do this is to bifurcate disgustingness in terms of the two families of responses: the affect program responses and the core responses. We would then have DISGUSTINGNESS$_1$ and DISGUSTINGNESS$_2$, where the former is a disposition to produce affect-program responses and the latter is a disposition to produce core-disgust responses."
>
> (Goldman 2007, page 7)

6.3.3.1 Objections to Goldman's Role Functionalist Account

Though perhaps more appealing that filler accounts, role functionalism is not without its own difficulties. First, individual and cultural variation in disgust elicitors raises some low-grade worries for any dispositionalist account. Consider deep fried Twinkies. Some people, like my friend Smitty, find these absolutely delicious (really!);

others, like myself, find them completely and utterly disgusting. In the form of a deep fried Twinkie, then, we have a single entity that both bears the disposition to elicit a disgust response, and does not bear the disposition to elicit a disgust response.

Such surface contradictions can be overcome with fairly standard philosophic means, relativizing the disposition in question to particular individuals or groups of people. In so doing, however, we lose the elegance of the account with a massive proliferation of properties. There is no longer a single property of disgustingess, but as many properties as there are differently calibrated disgust responses: $\text{disgustingness}_{\text{Smitty}}$ $\text{disgustingness}_{\text{Dan}}$, and so on. We might choose one and argue that it is *the* property of disgustingness, but role functionalism in and of itself gives us few resources to do so. It would also be chauvinistic, and thus somewhat paradoxical, in light of the fact that the ability to avoid chauvinism is alleged to be one of the great virtues of functionalism.

Goldman's refined version of dispositionalism would exacerbate this proliferation of properties. Since it divides disgustingness into disgustingness_1 and disgustingness_2 based other considerations, adding the worries about variation would effectively double the number of properties dispositionalism would have to countenance: $\text{disgustingness}_{\text{Smitty1}}$ and $\text{disgustingness}_{\text{Smitty2}}$, $\text{disgustingness}_{\text{Dan1}}$ and $\text{disgustingness}_{\text{Dan2}}$, and so forth.

More worrisome than this embarrassment of riches is that Goldman's dispositionalist view, in bifurcating the property of disgustingness the way it does, makes distinctions that the disgust response does not. That disgust response exhibits a psychological unity. One of the foundational desiderata of the first half of this dissertation was to explain that integrity:

> *The Unity of the Response*: The characteristic disgust *response* is comprised of a number of distinct features. These features form a homeostatic cluster: they occur together as a package, and regardless of what triggers disgust on any particular occasion, once it is triggered the production of one element of the cluster is regularly accompanied by the production of the others. What accounts for the clustering of this idiosyncratic set of features? Why have these particular cognitive, behavioral and physiological elements merged into a single, unified, and apparently universally human, response type?

Indeed, the Entanglement thesis separated the components of that response into those that originated in a poison mechanism, and those that originated in a parasite mechanism. It also held that in modern humans, those two, initially distinctly operating and still distinguishable mechanisms became functionally integrated with each other, to the point that they form a single response comprised of elements of each. When that response is triggered, *all* of those elements are produced. The upshot of this is that a disposition to produce the affect-program responses will yield the entire disgust response, the affect-program responses *and* core-disgust responses. Likewise, a disposition to produce the core-disgust responses will also yield the entire disgust response, core-disgust responses and affect-program responses both. Because of the psychological unity of the response, a disposition to trigger one sub-mechanism is thereby a disposition to trigger them both. If the Entanglement thesis is correct, then on Goldman's account, the terms "disgustingness$_1$" and "disgustingness$_2$" would be coextensive.

Since he makes the suggestion in a revisionist spirit, these appeals to empirical facts might not be so damaging to his account. In my view, separating out disgustingness$_1$ and disgustingness$_2$, however, glosses over some of the most interesting features of the response and its corresponding property, including properties that suggest it is projected onto the world, when "projected" is understood properly. Moreover, it

may seem redundant. There are already words in the vernacular that capture roughly what the properties disgustingness$_1$ and disgustingness$_2$ aim to capture; they are "poisonous" and "infectious", respectively.

6.4 Sentimentalism and Fittingness Accounts

Coming at the property of disgustingness from a slightly different angle are metaethicists working in the Humean sentimentalist tradition. Their first concern is not with the mind/body problem or the relationship of the mental to the physical, but with the nature of normativity, the metaphysics of morality and value, and the content of evaluative thought. What unites sentimentalists qua sentimentalists is a conviction that evaluative judgments, including moral and aesthetic judgments, depend, somehow, on human sentiments or emotional responses. In the Humean language, to judge something to be morally good, or beautiful, is to have sentiments of approbation towards it. The differences between sentimentalist views can often be traced to differences in the way each view understands the nature of the connection between sentiments and evaluative judgments.

6.4.1 Metaethical Concerns and Functionalist Accounts

Because metaethicists approach these issues from a quite different direction than philosophers of mind or mainstream analytic metaphysicians, it will be useful to quickly point out why they might reject either variety of functionalist account of disgustingness out of hand.

Metaethicists, and sentimentalists in particular, see in disgustingness a property that can serve as something of a simpler model for developing their views about more general issues about moral properties and concepts. The idea is that judgments about

disgustingness, discourse about disgustingness, arguments about disgustingness, all behave in ways that are very similar to judgments, discourse, and arguments about, for instance, whether some action is morally good. By gaining a better understanding of the nature of disgust and disgustingness, we can make some progress in understanding the nature of moral judgment and morality.

It turns out, then, that metaethicists are interested in disgustingness exactly where it differs from color. Unlike in ascriptions and judgments of color, for instance, there appears to be a role for reasoning to play in our assessments of disgustingness. In McDowell's terminology, such ascriptions and judgments of disgustingness are located in the "space of reasons".

Luckily, we can elaborate on this prevalent but somewhat obscure terminology. Judgments of disgustingness have the property of being, in David Wiggins' terminology, *essentially contestable*: they are subject to rational criticism and debate, perhaps ineluctably so. Those judgments are also *interpersonally authoritative:* they appear to carry the implication that others should agree with them, on pain of error or irrationality. Culinary debates over such alleged delicacies as escargot, fried locusts and deep fried Twinkies, or conflicting attitudes over the moral status of homosexuality all suggest that disgustingness and judgments about disgustingness, unlike perceptions of color, have these properties. The point is also made vivid by D'Arms and Jacobson's discussion of an American classic:

> "Consider the heretical view that the quintessential American delicacy, the Big Mac, is disgusting. A dispositionalist might try pointing out how many billions of them have been sold worldwide, but that would be to no avail. The heretic does not doubt that most people love them; it's just that this fails to move her. Evidently, most people's taste is abominable. Just look at the thing, she might add, all fatty and processed in its cardboard box, dripping with "special sauce." If

> you don't see what is disgusting about that, so much the worse for you."
> (D'Arms and Jacobson 2000, page 727)

Functionalist accounts of disgustingness have difficulty accommodating the variation and essential contestability of judgments of disgustingness, and the fact that they can be subject to reasoning and debate in this particular way. One might point out to a colorblind person that their color perception is incorrect, but this would not be a criticism of their evaluative or reasoning capacities, and it is unlikely that they would hold their judgments of color perception to be authoritative. Unlike the case of color, debates about these types of judgments have been construed as debates about when particular responses are *appropriate*, as we will see.[53]

6.4.2 Fittingness

Much work has been done, especially in the last 50 years or so, in developing sentimentalism. Here we will consider in more detail at a proposal by D'Arms and Jacobson, which I'll call their fittingness account. This isn't completely arbitrary: D'Arms and Jacobson take themselves to be acting in an ecumenical spirit, working with assumptions common to all current, neosentimentalist positions. What, then, is "fittingness", and what work is the notion supposed to do? Some brief background will help. In a useful overview of the development of sentimentalism in the 20th century, Nichols (2004, Chapter 4) shows how criticisms of traditional forms of sentimentalism (emotivist theories that construe moral claims as expressing occurant emotions, for instance) coalesced into a set of constraints on future sentimentalist theories. In

[53] I do not mean to suggest that I agree that explaining essential contestability and interpersonal authoritativeness should be seen as a condition of adequacy, or that any serious account of disgustingness (or an other property) should be able to capture those properties of the relevant discourse. I am merely pointing out the types of concerns that drive theorists who have a primarily metaethical agenda.

particular, those working in the tradition shared a loose consensus about three features of moral judgment that any sentimentalist view must account for:

1. Emotion plays a crucial role in moral judgment

2. A person can judge something wrong even if he has lost all feelings about it

3. Reasoning plays a crucial role in moral judgment

Together, these were widely taken to be conditions of adequacy on any viable sentimentalist account. What has come to be called neosentimentalism (Blackburn 1998, Gibbard 1991, Wiggins 1991) emerged from attempts to satisfy these three constraints.

According to D'Arms and Jacobson (2000), the controlling idea of neosentimentalism is that evaluative judgments, or at least an important subset of evaluative judgments, are best understood as judgments about the appropriateness of a particular emotional response. In their own words:

> "The crucial idea, which we take to be the defining characteristic of neosentimentalism, is that an important set of evaluative concepts (or terms or properties) is best understood as invoking a normative assessment of the appropriateness (or merit or rationality) of some associated emotional response. Hence,
>
> > (RDT) To think that X has some evaluative property Φ is to think it appropriate to feel F in response to X.
>
> For the neosentimentalist, to think a sentiment appropriate in the relevant sense is a normative judgment … in favor of feeling it"
>
> (D'Arms & Jacobson 2000, 729).

The main innovation over traditional sentimentalism is the emphasis on the *appropriateness* of an emotional response, rather than on the emotional response itself. On this view, and contrary to simple emotivist views, one need not actually experience any emotion when judging that some action is morally wrong - one might merely judge that it would be appropriate to feel guilt if one engaged in that action (Gibbard 1991).

This satisfies constraint 2, and of course constraint 1 is satisfied because the judgment is about the appropriateness of an emotional response. Judgments about whether or not guilt actually is appropriate can be subject to rational criticism and debate, thus satisfying constraint 3 as well.

Despite satisfying these three constraints, though, D'Arms and Jacobson (2005) argue that neosentimentalist theories constructed around the core idea of the RDT are still inadequate. D'Arms and Jacobson's fittingness account is a proposal for patching up the RDT, which they argue is ambiguous as it stands. More specifically, they claim that "appropriate" is too unconstrained to account for the response-dependent properties that the RDT is supposed to capture. There are many different ways in which an emotional response might be appropriate or not, and only some of them have to do with whether or not the object X actually has the property. For instance, it might be morally inappropriate to be disgusted by someone suffering from leprosy, but this is ancillary to the question of whether or not a leper is genuinely disgusting. Analogously, it might be morally inappropriate to laugh at a clever, but racist joke. But the inappropriateness of amusement or laughter is likewise a different issue from whether the joke actually has the response dependent property of funniness.

In order to resolve the problem, D'Arms and Jacobson offer their fittingness account, which is inspired by passages that can be found in earlier work by sentimentalists like Brandt (1946) and Wiggins (1987). This account is supposed to disambiguate the crucial notion of appropriateness. Judgments of fit are a particular type of judgments of appropriateness, but considerations of fit are neither considerations of the moral, prudential, or strategic appropriateness of the relevant response. Rather, to judge

an emotion a fitting response to some object or action is to judge that the object or action has the corresponding property, that the emotion is functioning as it was made to, and that the emotion presents the object correctly. Judgments of fittingness are taken to be anchored in the structure and function of the emotions themselves: "Assessments of fittingness are attempts to make sense of or criticize our emotions using standards that speak to the distinctive concerns we take them to embody" (D'Arms 2005, page 11). Wiggins puts the point, another way: ''we can fix on a response . . . and then argue about what the marks are of the property that the response itself is made for" (1987). In general,

> "[C]onsiderations of fittingness are all and only those considerations about whether to feel shame, amusement, fear, and so forth bear on whether the emotion's evaluation of the circumstance gets it right: whether the situation really is shameful, funny, fearsome, and so forth. Norms of fittingness are one kind of rational norm for appraising emotional responses – albeit an especially important and effective kind – and they must be distinguished from other forms of appraisal (D'Arms & Jacobson 2003, page 132).

The fittingness account can be recast in similar terms to the RDT it is supposed to supplant:

> (FRDT): To think that X has some evaluative property Φ is to think it *fitting* to feel F in response to X.

A more concrete example can help illustrate the point. We might wonder, for instance, whether or not garden slugs are really disgusting. According to the fittingness account the question can be rephrased thus: is the emotion of disgust a fitting response? Claiming that garden slugs have the property of disgustingness is to think it fitting to feel disgust in response to garden slugs. Garden slugs are really disgusting if and only if disgust is fitting response to them. Likewise, laughter and amusement might be a fitting

response to a very clever racist joke, even if it would be inappropriate to laugh or be amused on moral grounds.

6.4.2.1 Objections to Fittingness Accounts

As we have seen, the notion of fittingness is supposed to bolster the neosentimentalist characterization of evaluative judgments, and of the response dependent properties they purport to be about. More specifically, it is supposed to, to a first approximation, provide a way to supplement the RDT in such as way that we can draw a principled distinction between things that are really disgusting, i.e. things towards which it is fitting to feel disgust, and things that are not really disgusting, i.e. even those things one might have good moral, prudential, strategic reason to feel disgusted by. In this, fittingness is supposed to yield an account of the property of disgustingness, and thus locate its place in nature.

Given its wide scope and lofty ambitions, neosentimentalism can be difficult to wrap one's mind around and evaluate. I, for one, find myself continually frustrated by the seemingly ubiquitous but rudimentary mistake of confusing concepts and properties, and am often left with a vague sense that explanation and justification are being conflated somewhere, even if it is difficult to pinpoint where the mistake is made. Other authors (Griffiths forthcoming) have argued from premises about the social, strategic, and Machiavellian character of many emotions to the conclusion that despite the hopes of sentimentalists, that there is no easy route from the emotions to normativity. Despite my own earlier enthusiasm for the tradition, I now suspect some such conclusion is correct.

In the local case of disgustingness, however, we can mount an objection that goes straight to the heart of the fittingness account. As spelled out in the previous chapter,

there is nearly always an *imperfect fit* between the full disgust response and the entities that trigger it. In fact, with disgust the imperfect fit can be traced to three different forces. The first is a hair trigger activation and "better safe than sorry" logic, which yields a high number of false positives. The second and third stem from the Entanglement and Co-opt theses. The disgust response is comprised of a set of features that nomologically cluster, but elements of those features can be traced back to two different systems, one pertaining to parasites, the other pertaining to poisons. As such, it presents things as having a cluster of properties, including offensiveness, contamination, nauseatingness. However, since disgust is a kludge, very few things will actually have all of the properties that the disgust response presents them as having, even in the cases where it is performing one of its primary functions. Moreover, in cases where it is performing an auxiliary function, where it has been co-opted and put to uses for which it did not initially evolve, the response is even less fitted to its object. It follows that disgust is a fitting response to almost nothing, and thus, by the lights of the fittingness account, that nothing will really be disgusting.

For those who still reject outright eliminativism and wish to vindicate or even merely make sense of the common sense, discourse, and any social practices having to do with disgustingness, this may seem to leave us left with only projectivism. While I see no problem with that option, it is unlikely that D'Arms and Jacobson, the two most vocal proponents of fittingness accounts, will be similarly sanguine. They have explicitly argued against projectivist views, and present their own as an alternative to them (D'Arms and Jacobson 2005). Granted, the targets of their arguments are previous

versions of projectivism; it is not yet clear how they would react to the psychologized version I favor.

6.7 Conclusion

In this chapter we have examined and criticized two account of disgustingness that derive from functionalist views in the philosophy of mind and accounts of color, and one account that derives from sentimentalist views in metaethics. Many of my objections stem from the inability of these accounts to accommodate certain facts about disgust that were established in the first half of the dissertation – that it is a piecemeal conglomeration of different mechanisms, that it has acquired multiple functions generating even greater the lack of fit between the response and the auxiliary functions it was coopted to perform, and that due to a flexible acquisition system, it allows for substantial variation.

I certainly do not claim to have decisively refuted any of the accounts considered above, and other philosophers may find ingenious ways to supplement or extend those accounts so they can better deal with the objections levied against them. Mainly I have been concerned to point out the weaknesses and shortcomings of their current forms. In some cases, I have suggested why projectivism, or a psychologized projectivism, does a much better job of accounting for the property of disgustingness and locating it in the natural order. I suspect that the same will be true of a variety of other properties that are difficult to locate in nature, but this is not the place to argue for that claim.

Bibliography

Appiah, K. A. (1995). The uncompleted argument: Du Bois and the illusion of race. In L. A. Bell and D. Blumenfeld (Eds.), *Overcoming racism and sexism* (pp. 59-78). Lanham, MD: Rowman and Littlefield.

Ariew, A. (2003). Ernst Mayr's 'ultimate/proximate' distinction reconsidered and restructured. *Biology and Philosophy*, 18(4), 553-565.

Armstrong, D. (1968). *A materialist theory of mind*. London: Routledge & Kegan Paul.

Atran, Scott. (2002). *In god we trust: The evolutionary landscape of religion*. New York: Oxford University Press.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, III, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117-150.) Washington, DC: American Psychological Association.

Bandura, A. (1992). Social cognitive theory of social referencing. In S. Feinman (Ed.), *Social referencing and the social construction of reality in infancy* (pp. 175-208). New York: Plenum.

Barkow, J., Cosmides, L., & J. Tooby. (1992). *The adapted mind*. New York: Oxford University Press.

Baron-Cohen, S. 1995. *Mindblindness*. Cambridge, MA: The MIT Press.

Barth, F. (Ed.) (1969). *Ethnic groups and boundaries: The social organization of cultural differences*. Boston: Little Brown & Co.

Becker, E. (1973). *The denial of death*. New York: Simon & Shuster.

Bennett, K. 2007. Mental causation. *Philosophy Compass*, 2(2), 316-337.

Bernstein, I. (1999). Taste aversion learning: A contemporary perspective. *Nutrition*, 15(3), 229-234.

Blackburn, S. (1984). *Spreading the word*. New York: Oxford University Press.

Blackburn, S. (1993). *Essays in quasi-realism*. New York: Oxford University Press.

Blackburn, S. (2000). Critical notice of Frank Jackson, *From metaphysics to ethics: A defense of conceptual analysis*. *Australasian Journal of Philosophy*, 78(1), 119-24.

Block, N. (1980). Troubles with functionalism. In W. Savage (Ed.), *Minnesota Studies in the Philosophy of Science, Vol IX: Perception and Cognition* (pp. 261-325). Minneapolis: University of Minnesota Press.

Bloom, P. (2004). *Descartes' baby*. New York: Basic Books.

Borg, J., Lieberman, D., & K. Kiehl. (2006). Pathogen, sexual, and moral disgust – distinct or common neural networks?" Poster.

Bowles, S. and Gintis, H. (1998). The moral economy of community: Structured populations and the evolution of prosocial norms. *Evolution and Human Behavior*, 19, 3-25.

Bowles, S. and Gintis, H. (2001). Community governance. In A. Nicita and U. Pagano, (Eds.), *The evolution of economic diversity*. London: Routledge.

Boyd, Richard. (1991). Realism, anti-foundationalism and the enthusiasm for natural dinds. *Philosophical Studies,* 61, 127-148.

Boyd, R. & P. J. Richerson. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Boyd, R. & P. J. Richerson. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethnology and Sociobiology* 13, 171–95.
Boyd R., & P. Richerson. (1996). Why culture is common, but cultural evolution is rare. *Proceedings of the British Academy*, 88,77–93.
Boyd R., & P. Richerson. (2005). *The Origin and Evolution of Cultures*. New York: Oxford University Press.
Boyer, P. (2001). *Religion explained: The evolutionary origins of religious thought*. New York: Basic Books.
Brandt, R. (1946). Moral valuation. *Ethics*, 56, 106-21.
Byrne, A. and Hilbert, D. (2003). Color realism and color science. *Behavioral and Brain Sciences*, 26(1), 3-64.
Chalmers, D. (2002). *The philosophy of mind*. New York: Oxford.
Chalmers, D. (2003). Consciousness and it's place in nature. In Stich, S. & F. Warfield (Eds.), *Blackwell guide to philosophy of mind*. New York: Blackwell.
Charash, M., & McKay, D. (2002). Attention bias for disgust. *Journal of Anxiety Disorders,* 16(5), 529-541.
Churchland, P. M. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78, 67-9
Coan, J. & Allen, J. (2003). Varieties of emotional experience during voluntary emotional facial expression. *Ann. N.Y. Acad. Sci.* 1000, 375–379.
Cohen, J. (Draft). It's not easy being green: Hardin and color relationalism.
Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to 'prejudice.' *Journal of Personality and Social Psychology,* 88, 770-789.
Cummins, R. (1983). *The nature of psychological explanation*. Cambridge, MA: The MIT Press.
Cummins, R. (2000). 'How does it work?' vs. 'What are the laws?': Two conceptions of psychological explanation. In F. Keil and R. Wilson (Eds), *Explanation and cognition* (pp 117-145). Cambridge, MA: The MIT Press.
Curtis, V., Aunger, R. & Rabie T. (2004). Evidence that disgust evolved to protect from risk of disease. *Proceedings of the Royal Society: Biological Science Series B*, 271(4), S131-S133.
Curtis, V & Biran, A. (2001). Dirt, disgust, and disease: Is hygiene in our genes? *Perspectives in Biology and Medicine*, 44(1), 17-31.
D'Arms, J & D. Jacobson. (2000). Sentiment and value. *Ethics*, 110, 722-48.
D'Arms, J. & D. Jacobson. (2003). The significance of recalcitrant emotions. In A. Hatzimoysis, ed. *Philosophy and the emotions*, Cambridge, Cambridge University Press.
D'Arms, J. & D. Jacobson. (2005). Sensibility theory and projectivism. In D. Copp (Ed.) *The Oxford handbook of ethical theory*. Oxford: Oxford University Press.
Daly, M. & Wilson, M. (1988). *Homicide*. Hawthorne, NY: Aldine.
Darwall, S., Gibbard, A. & P. Railton. (1993). Toward fin de siecle ethics: Some trends. *The Philosophical Review*, 101(1), 115-189.
Darwin, C. (1872). *The expressions of emotions in man & animals* (1st ed.). New York: Philosophical Library.

Davey, G., Forster, L, & G. Mayhew. (1993). Familial resemblances in disgust sensitivity and animal phobias. *Behavior Research and Therapy*, 31(1), 41-50.

De Sousa, R. (1987). *The rationality of emotion*. Cambridge: The MIT Press.

De Caro, M. & D. Macarthur. (2004). *Naturalism in question*. Cambridge, MA: Harvard University Press.

Dennett, D. (1978). *Brainstorms*. Montgomery, VT: Bradford Books.

Dennett, D. (1981). True believers: The intentional strategy and why it works. In Haugeland, J. (Ed.) *Mind Design II*. Oxford: Oxford University Press.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown, and Co.

Dennett, D. (1991). Real patterns. *Journal of Philosophy*, 88(1), 27-55.

Dennett, D. (1995). *Darwin's dangerous idea*. New York: Simon and Schuster.

Dennett, D. 2006. *Breaking the spell: Religion as natural phenomenon*. New York: Penguin Publishers.

Doris, J. & Stich, S. (2005). As a matter of fact: Empirical perspectives on ethics. In F. Jackson and M. Smith (Eds.), *The Oxford handbook of contemporary philosophy*. Oxford: Oxford University Press.

Doris, J. & Stich, S. (2006). Moral psychology: Empirical approaches. *The Stanford Encyclopedia of Philosophy (Summer 2006 Edition)*, Edward N. Zalta (Ed.), URL = http://plato.stanford.edu/archives/sum2006/entries/moral-psych-emp/.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169-200.

Ekman, P. (2003). Darwin, deceit, and facial recognition. *Ann. N.Y. Acad. Sci.* 1000, 205-221.

Ekman, P. & Davidson, R. (1993). Voluntary smiling changes regional brain activity. *Psychological Science*, 4(5), 342-5.

Fallon, A. E., & Rozin, P. (1983). The psychological bases of food rejections by humans. *Ecology of Food and Nutrition*, 13, 15-26.

Fallon, A., Rozin, R., & R. Pliner. (1984). The child's conception of food: The development of food rejections with special reference to disgust and contamination sensitivity. *Child Development*, 55, 566-575.

Faulkner, J., Schaller, M., Park, J., & L. Duncan. (2004). Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Group Processes & Intergroup Relations*, 7(4), 333-353.

Fessler, D., Arguello, A., Mekdara, J. & R. Macias. (2003). Disgust sensitivity and meat consumption: a test of an emotivist account of moral vegetarianism. *Appetite*, 41, 31-41.

Fessler, D., Eng, S. & Navarrete, C. (2005). Elevated disgust sensitivity in the first trimester of pregnancy: Evidence supporting the compensatory prophylaxis hypothesis. *Evolution and Human Behavior*, 26, 344-351.

Fessler, D. & Navarrete, C. (2003). Domain specific variation in disgust sensitivity across the menstrual cycle. *Evolution and Human Behavior*, 24, 406-417.

Fessler, D. & Navarrete, C. (2003). Meat is good to taboo: Dietary proscriptions as a product of the interaction of psychological mechanisms and social processes. *The Journal of Cognition and Culture*, 3(1),1-40.

Fessler, D. & Navarrete, C. (2004). Third-party attitudes toward sibling incest: Evidence for Westermarck's hypothesis. *Evolution and Human Behavior*, 25, 277-294.

Fessler, D., Pillsworth, E. & Flamson, T. (2004). Angry men and disgusted women: an evolutionary approach to the influence of emotion on risk taking. *Organizational Behavior and Human Decision Processes*, 95, 107–123.

Flanagan, O. (1991). *Varieties of moral personality: Ethics and psychological realism*. Cambridge, MA: Harvard University Press.

Fodor, J. (1968). *Psychological explanation*. New York, NY: Random House.

Fodor, J. (1975). *The language of thought*, Cambridge, MA: Harvard University Press.

Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: The MIT Press.

Frank, R. (1988). *Passions within reason: The strategic role of the emotions*. New York: W.W. Norton & Company.

Fukuyama, F. (1992). *The end of history and the last man*. New York: Harper Perennial.

Garcia, J., Hankins, W. & K. Rusiniak. (1974). Behavioral regulation of the milieu interne in man and rat. *Science*, 185, 824-831.

Geach, P. (1965). Assertion. *Philosophical Review* 74(4), 449-465.

Gelman, S. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment*. Cambridge, MA: Harvard University Press.

Gil-White, F. (2001). Are ethnic groups biological 'species' to the human brain? *Current Anthropology*, 42 (4), 515-554.

Gil-White, F. (Manuscript a). The study of ethnicity needs better categories.

Gil-White, F. (Manuscript b). Is ethnocentrism adaptive?

Goldman, A. (2007) A program for "naturalizing" metaphysics, with application to the ontology of events. *The Monist*.

Gould, S. J. & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings Of The Royal Society of London, Series B,* 205 (1161), 581-598.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science,* 6, 517-523.

Greenwald, A., McGhee, D. & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480.

Greenwald, A., Nosek, B. & Banaji, R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85,197-216.

Griffiths, P. (1997). *What the emotions really are*. Chicago: University of Chicago Press.

Griffiths, P. (2001). Emotion and expression. *International Encyclopedia of the Social and Behavioral Sciences*. Pergamon/Elsevier Science.

Griffiths, P. (2003). Basic emotions, complex emotions, Machiavellian emotions. In *Philosophy and the emotions,* (Ed.) A. Hatzimoysis. New York: Cambridge University Press.

Griffiths, P. (Forthcoming). "Ask not what your emotions can do for you…" – Emotions, normativity, and Machiavellian intelligence. Based on a talk given at ISRE Conference in Atlanta, GA, August 6th 2006.

Guthrie, S. (1993). *Faces in the clouds: A new theory of religion*. New York: Oxford University Press.

Haidt, J. (2006). *The happiness hypothesis: Finding modern truth in ancient wisdom*, New York: Basic Books.

Haidt, J . (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*. 108, 814-834.

Haidt, J., & Hersh, M. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology,* 31, 191-221.

Haidt, J., Killer, S. & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65(4), 613-628.

Haidt, J., C. McCauley, and P. Rozin (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences*, 16, 701-713.

Haidt, J., Rozin, P., McCauley, C. & Imada, S. (1997). Body, psyche, and culture: The relationship between disgust and morality. *Psychology and Developing Societies*, 9, 107-131.

Hardin, C. (1988). *Color for philosophers: Unweaving the rainbow*. Indianapolis: Hackett Publishing.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuro-imaging responses to extreme outgroups. *Psychological Science,* 14(10), 847-853.

Hatfield, E., J. Cacioppo & R. Rapson. (1994). *Emotional contagion*. New York: Cambridge University Press.

Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: HarperCollins.

Heath, C., Bell, C., & Sternberg, E. (2001). Emotional selection in memes: The case of urban legends. *Journal of Personality & Social Psychology*, 81, 1028-1041.

Hejmadi, A., Rozin, P., & Siegal, M. (2004). Once in contact, always in contact: Contagious essence and conceptions of purification in American and Hindu Indian children. *Developmental Psychology*, 40(4), 467-476.

Henrich et al. (2005). 'Economic man' in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral & Brain Sciences*, 28, 795-855.

Henrich, J. & Gil-White, F. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22,165-196.

Henrich, J. & Boyd, R. (1998). The evolution of conformist transmission and the emergence of between group differences. *Evolution and Human Behavior*, 19, 215-241.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology*, 12, 123-135.

Heyes, C. M. (1993). Imitation, culture and cognition. *Animal Behaviour*, 46, 999–1010.

Hume, D. (1975). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. P. Nidditch (Ed.). Oxford: Clarendon Press.

Hume, D. (1978). *A Treatise of Human Nature*. . P. Nidditch (Ed.). Oxford: Clarendon Press.

Jackson, F. (1996). Mental causation. *Mind*, 105, 377-413.

Kass. L. (2002). *Life, liberty, and the defense of dignity: The challenge to bioethics*. New York: Encounter Books.

Kelly, D., Machery, E. & Mallon, R. (Forthcoming). Racial cognition and normative racial theory. In J. Doris, W.S. Armstrong, S. Nichols, & S. Stich (Eds.), *The handbook of moral psychology*.

Kelly, D., Machery, E., Mallon, R., Mason, K., & Stich, S. (2006). The role of psychology in the study of culture. *Behavioral and Brain Sciences*, 29(4), 355.

Kelly, D., Stich, S., Haley, K., Eng, S. & Fessler, D. (2007). Harm, affect and the moral / conventional distinction. *Mind and Language*, 22(2), 117-131.

Kelly, D. & Stich, S. (Forthcoming). Two theories about the cognitive architecture underlying morality. In P. Carruthers, S. Laurence & S. Stich (Eds.), *Innateness and the structure of the mind: Foundations and the future*. (New York: Oxford University Press).

Keltner, D. & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, 13, 505-521.

Kim, J. & Sosa, E. (1999). *Metaphysics: An anthology*. New York: Blackwell Publishing.

Knapp, C. (2003). Demoralizing disgustingness. *Philosophy and Phenomenological Research*, 66 (2), 253-276.

Knobe, J. (Forthcoming). "Reason explanation in folk psychology." *Midwest Studies in Philosophy*.

Krolak-Salmon, P., Henaff, M.A., Isnard, J., Tallon-Baudry, C., Guenot, M., Vighetto, A., Bertrand, O., and Mauguiere, F. (2003). An attention modulated response to disgust in human ventral anterior insula. *Ann. Neurol.* 53, 446–453.

Kurzban, R., J. Tooby and L. Cosmides. (2001). Can race be erased? Coalitional computation and social categorization. *Proceeding of the National Academy of Science*, 98(26), 15387-15392.

Kurzban, R. & Leary, M. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin*. 127(2), 187-208.

Leakey, R. (1994). *The origin of humankind*. New York: Basic Books.

Lerner, J., Small, D., & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science*, 15(5), 337-341.

Leslie, A. (1987). Pretence and representation: The origins of "theory of mind," *Psychological Review* 94, 412-426.

Levenson, R. (1992). Autonomic nervous system differences among emotions. *Psychological Science*, 3(1), 23-27.

Lewis, D. 1966. An argument for the identity theory. *Journal of Philosophy* 63, 17-25. Reprinted (with additions) in David Rosenthal (Ed.) *Materialism and the mind-body problem* (1971; Prentice-Hall) and in his *Philosophical papers I*.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3), 249-58.

Machery, E. & Stich. S. (Forthcoming). The evolution of morality. In J. Doris, W.S. Armstrong, S. Nichols, & S. Stich (Eds.), *The handbook of moral psychology*.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York: Penguin Books.

Marcus, G. (Forthcoming). *Kluge: The haphazard construction of the human mind*.

Mason, W. 1985). Experiential influences on the development of expressive behaviors in rhesus monkeys. In G. Zivin (Ed.), *The development of expressive behavior: Biology-environmental interactions*. New York: Academic Press.

Mayr, E. (1960). The emergence of evolutionary novelties. In Tax, S. (Ed.), *Evolution after Darwin: Vol. 1, the evolution of life*. Chicago, University of Chicago Press.

Mayr, E. (1961). Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science,* 134, 1501-1506.

McDowell, J. (1985). Values and secondary qualities. Reprinted in McDowell, 1998, 131-150.

McDowell, J. (1987). Projection and truth in ethics. Reprinted in McDowell, 1998, 151-166.

McDowell, J. (1998). *Mind, Value, and Reality*. Cambridge: Harvard University Press.

McElreath, R., Boyd, R. & Richerson, P. (2003). Shared norms can lead to the evolution of ethnic markers. *Current Anthropology*, 44(1), 123–29.

McElreath et al. (2005). Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior*, 26, 483-508.

McLaughlin, B. (2003). The place of color in nature. In R. Mausfeld & D. Heyer (Eds.), *Colour: Connecting the mind to the physical world* (pp. 475-505). Oxford: Oxford University Press.

McLaughlin, B. (Forthcoming). Type materialism for phenomenal consciousness. In *The Blackwell Companion to Consciousness*. New York: Blackwell Publishers.

Medin, D. & Atran, S. (Eds.) (1999). *Folkbiology*. Cambridge, MA: The MIT Press.

Miller, S. (1986). Disgust: Conceptualization, development and dynamics. *International Review of Psychoanalysis*, 13, 295-307.

Miller, S. (1993). Disgust reactions: There determinants and manifestations in treatment. *Contemporary Psychoanalysis*, 29(4), 711-735.

Miller, W. (1997). *The anatomy of disgust*. Harvard University Press: Cambridge, MA.

Morris, P., Doe, C. & Godsell, E. (2007). Secondary emotions in non-primate species? Behavioral reports and subjective claims by animal owners. *Cognition and emotion*.

Murphy, S., Haidt, J. & Bjorkland, F. (Manuscript-2000). Moral dumbfounding: When intuition finds no reason.

Nabi, R. (2002). The theoretical versus the lay meaning of disgust: Implication for emotion research. *Cognition & Emotion*, 16(5), 695-703.

Nado, J., Kelly, D. & Stich, S. (Forthcoming). Moral judgment. In J. Symons & P. Calvo (Eds.) *The Routledge Companion to the Philosophy of Psychology*. New York: Routledge.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435-50.

Nemeroff, C. & Rozin, P. (2000). The makings of the magical mind: The nature and function of sympathetic magical thinking. In K. Rosengren, C. Johnson & P. Harris (Eds.), *Imagining the Impossible*. Cambridge University Press: New York.

Nichols, S. (2001). Innateness and moral psychology. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The innate mind: Structure and content*. New York: Oxford University Press.

Nichols, S. (2002a). Norms with feeling: Towards a psychological account of moral judgment. *Cognition,* 84, 221-236.

Nichols, S. (2002b). On the genealogy of norms: a case for the role of emotion in cultural evolution. *Philosophy of Science*, 69, 234-255.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. New York: Oxford University Press.
Nichols, S. and Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, 100(3), 530-542.
Nichols, S. and Mallon, R. (Forthcoming). Rules and moral psychology. In J. Doris, W.S. Armstrong, S. Nichols, & S. Stich (Eds.), *The handbook of moral psychology*.
Noe, A. (2004). *Action in Perception*. Cambridge, MA: The MIT Press.
Nucci, L. (2001). *Education in the moral domain*. Cambridge: Cambridge University Press.
Nussbaum, M. (2004). *Hiding from humanity: Disgust, shame, and the law*. Princeton, NJ: Princeton University Press.
Park, J., Faulkner, J., & Schaller, M. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with physical disabilities. *Journal of Nonverbal Behavior*, 27(2), 65-87.
Penfield, W., and Faulk, M.E. (1955). The insula: Further observations on its function," *Brain* 78, 445–470.
Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C. R., Gray, J. A., David, A. S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, 389(6650), 495-498.
Pinker, S. (1997). *How the mind works*. New York: W.W. Norton & Co.
Place, U.T. (1956). Is Consciousness a Brain Process? *British Journal of Psychology*, 47, 44-50.
Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. New York: Oxford University Press.
Prinz, J. Forthcoming. *The emotional construction of morals*.
Putnam, H. (1967). Psychological predicates. In W. H. Capitan and D. D. Merrill (Eds.), *Art, mind and religion*. Pittsburgh, Penn.: University of Pittsburgh Press.
Putnam, H. (1968). Brains and behavior. Reprinted in Putnam, 1975, 325-41.
Putnam, H. (1975). *Mind, language and reality. Philosophical papers, vol. 2*. Cambridge: Cambridge University Press.
Putnam, H. (1990). *Realism with a human face*, edited by Conant, J. Cambridge, MA: Harvard University Press.
Putnam, H. (1999). Pragmatic realism. In Kim, J. & E. Sosa (Eds.), *Metaphysics: An anthology*. New York: Blackwell Publishing.
Rachels, J. (2002). *The elements of moral philosophy*. New York: McGraw Hill.
Richerson, P. & Boyd, R. (1998). The evolution of human ultra-sociality. In I. Eibl-Eibisfeldt and F. Salter (Eds.), *Ideology, warfare, and indoctrinability*. Oxford: Oxford University Press.
Richerson, P. & Boyd, R. (1999). Complex societies: The evolutionary origins of a drude superorganism. *Human Nature*.
Richerson, P. & Boyd, R. (2000). Climate, culture, and the evolution of cognition. In . C. M. Heyes (Ed), *The evolution of cognition*. Cambridge, MA: The MIT Press.
Richerson, P. & Boyd, R. (2005). *Not by genes alone*. Chicago: University of Chicago Press.

Rozin, P. (1997). Moralization. In Brandt, A. & Rozin, P. (Eds.), *Morality + Health*. New York: Routledge.

Rozin, P. & Fallon, A. (1987). A perspective on disgust. *Psychological Review*, 94, 23-41.

Rozin, P., Fallon, A. & Augustoni-Ziskind, M. (1985). The child's conception of food: The development of contamination sensitivity to "disgusting" substances. *Developmental Psychology*, 21, 1075-79.

Rozin, P., Haidt, J., & McCauley, C. (2000). Disgust. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions, 2nd Edition*, New York: Guilford Press.

Rozin, P., Haidt, J., McCauley, C, and Imada, S. (1997). Disgust: Preadaptation and the cultural evolution of a food-based emotion. In H.M. Macbeth (Ed.), *Food preferences and taste: Continuity and change*. Oxford: Berghahn.

Rozin, P., Hammer, L., Oster, H., Horowitz, T., &and Marmora, V. (1986). The child's conception of food: Differentiation of categories of rejected substances in the 16 months to 5 year age range. *Appetite*, 7, 141-151.

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology,* 66(5), 870-881.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76(4), 574-586.

Rozin, P., Markwith, M. & Nemeroff, C. (1992). Magical contagion beliefs and fear of AIDS. *Journal of Applied Social Psychology*, 22(14), 1081-1092.

Rozin, P., Millman, L. & Nemeroff C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, 50, 703-712.

Rozin, P. & Nemeroff, C. (1990). The laws of sympathetic magic: A psychological analysis of similarity and contagion. In J. Stigler, R. Schweder, & G. Herdt, (Eds.), *Cultural psychology: Essays in comparative human development*, Cambridge: Cambridge University Press.

Rozin, P., Nemeroff, C., Horowitz, M., Gordon, B. & Voet, W. (1995). The borders of the self: Contamination sensitivity and potency of the body apertures and other body parts. *Journal of Research in Personality*, 29, 318-340.

Rozin, P. & Siegal, M. (2003). Vegemite as a marker of national identity. *Gastronomica – The Journal of Food and Culture*, 3(4), 63-67.

Ryle, G. 1949. *The concept of mind*. London: Hutchinson.

Sanfey, A., Rilling, J., Aronson, J., Nystrom, L. & Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755-1758.

Schnall, S., Haidt, J. & Clore, G. (Manuscript-2004). Disgust as embodied moral judgment.

Schiffer, S. (1981). Truth and the theory of content. In H. Parret and J. Bouvaresse (Eds.), *Meaning and understanding*. Berlin: Walter de Gruyter.

Shweder, R., Much, N., Mahapatra, M. and Park, L. (1997). The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In Brandt, A. & Rozin, P. (Eds.), *Morality + Health*. New York: Routledge.

Siegal, M. (1988). Children's knowledge of contagion and contamination as causes of illness. *Child Development*, 59, 1353-1359.

Siegal, M. & Share, D. (1990). Contamination sensitivity in young children. *Developmental Psychology*, 26(3), 455-458.

Small, D., Zald, D., Jones-Gotman, M., Zatorre, R., Pardo, J., Frey, S., & M. Petrides. (1999). Brain imaging: Human cortical gustatory areas: A review of functional neuroimaging data. *NeuroReport*, 10, 7-14.

Smart, J.J.C. (1959). Sensations and brain processes. *Philosophical Review*, 68, 141-156.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. New York: Blackwell Publishers.

Sripada, C. (2005). Punishment and the strategic structure of moral systems. *Biology and Philosophy*, 20, 767–789.

Sripada, C. (Forthcoming). Adaptationist and culturist explanations of human behavior of behavior. In P. Carruthers, S. Laurence & S. Stich (Eds.), *Innateness and the structure of the mind: Foundations and the future*. (New York: Oxford University Press).

Sripada, C. (Manuscript). Carving the social world at its joints: Moral norms and conventions as natural kinds.

Sripada, C. & S. Stich. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The innate mind: Culture and cognition*. New York: Oxford University Press.

Sterelny, K. (2003). *Thought in a hostile world*. New York, Blackwell.

Stich, S. (1983). *From folk psychology to cognitive science: The case against belief*. Cambridge, MA: The MIT Press.

Stich, S. (1996). Naturalism, positivism, and Puritanism. In *Deconstructing the Mind*. New York: Oxford University Press.

Stich S. (2006). Is morality an elegant machine or a kludge? *Journal of Cognition and Culture*, 6(1), 181-189.

Stich, S. & Laurence, S. (1994). Intentionality and naturalism. In French P A, Uehling, T E, Jr. (Eds.), *Midwest Studies in Philosophy: Naturalism*, 19,159-182

Stroud, B. (1996). The charm of naturalism. Reprinted in De Caro & Macarthur (Eds.), 2004, 21-35.

Tajfel, H., Flament,C., Billig, M.G., and Bundy, R.P. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1, 149-75.

Tetlock, P.E., Kristel, O., Elson, B., Green, M., and Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78, 853-870.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, 16, 495--552.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tooby, J. & L. Cosmides. (2005). Evolutionary psychology: Conceptual foundations. In D. Buss (Ed.), *The handbook of evolutionary psychology*. New York: Wiley.

Trivers R (1971). The evolution of reciprocal altruism. *Quarterly Journal of Biology*, 46, 35-57.

Turiel, E. (1983). *The development of social knowledge*. Cambridge: Cambridge University Press.

Turner, J.C. (1984). Social identification and psychological group formation. In H. Tajfel (Ed.), *The social dimension: European developments in social psychology, Vol 2*. Cambridge: Cambridge University Press.

Ware, J., Jain, K., Burgess, I., & Davey, G. (1994). Disease-avoidance model: Factor analysis of common animal fears. *Behavior Research and Therapy*, 32(1), 57-63.

Webb, K. & Davey, G. (1993). Disgust sensitivity and fear of animals: Effect of exposure to violent and revulsive material. *Anxiety, Coping and Stress*, 5, 329-335.

Wheatley, T. & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, 16(10), 780-784.

Wicker, B., Keysers, C., Plailly, J., Royet, J., Gallese, V. & Rizzolatti, G. (2003). Both of us disgusted in *my* insula: The common neural basis of seeing and feeling disgust. *Neuron* 40, 655-664.

Wiggins, D. (1976). Truth, invention, and the eaning of life. Reprinted in Wiggins, 1987b, 87-137.

Wiggins, D. (1987a). A sensible subjectivism? Reprinted in Wiggins, 1987b, 185-214.

Wiggins, D. (1987b). *Needs, values, truth*. Oxford: Basil Blackwell Ltd.

Wilson, T. (2002.) *Strangers to ourselves*. Cambridge, MA: Harvard University Press.

White, S. (2004). Subjectivity and the agential perspective. In M. De Caro & D. Macarthur (Eds.), *Naturalism in question*. Cambridge, MA: Harvard University Press.

Zhong, C. & K. Liljenquist. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, 313 (5792), 1451-2.

Curriculum Vita

Daniel Ryan Kelly

Education

1993-1997, Illinois Wesleyan University, Philosophy, English Literature, BA
1998-2001, Tufts University, Philosophy, MA
2001-2007, Rutgers University, PhD

Occupations: Teaching, Fellowships, and Professional Activities

Teaching Assistant, Mind & Language, Fall 1999 (Tufts)
Research Assistant, Center for Cognitive Studies, Tufts University, 1999-2000
Teaching Assistant, Philosophy & Film, Spring 2000 (Tufts)
Assistant Organizer of Rutgers-Princeton Graduate Student Conference, April 2002
Organizer of Rutgers-Princeton Graduate Student Conference, April 2003
Member and Contributor to the Moral Psychology Research Group, Fall 2003-present
Member and Contributor to the *Innate Mind* project, University of Maryland and University of Sheffield, Spring 2003-Summer 2004
Participant in the planning workshop for new international project on *Culture and the Mind*, London, UK, February 2005
Instructor, Introduction to Philosophy, Fall 2005, Fall 2007
Instructor, Introduction to Logic, Spring 2006
Mind and Culture Fellowship, Rutgers Center for Cultural Analysis, 2006-2007
Referee for *Philosophical Explorations* May 2007-present
Referee for *Mind and Language*, September 2007-present

Publications

'Naturalization of Intentionality,' *The Springer Encyclopedic Reference of Neuroscience*, Ed. Michael Bilic, 2005. [with Kelby Mason and Dennis Whitcomb]
'Moral Realism and Cross-cultural Normative Diversity,' *Behavioral and Brain Sciences*, 2005, Vol 28(6): 830. [with Edouard Machery and Stephen Stich]
'The Role of Psychology in the Study of Culture,' *Behavioral and Brain Sciences*, 2006, Vol 29(4): 355. [with Edouard Machery, Ron Mallon, Kelby Mason, and Stephen Stich]
'Harm, Affect and the Moral / Conventional Distinction,' *Mind & Language*, 2007, Vol. 22 (2): 117-131. [with Stephen Stich, Daniel M. T. Fessler, Kevin J. Haley, and Serena Eng]
'Two Theories of the Cognitive Architecture Underlying Morality,' to appear in *The Innate Mind Vol 3.: Foundations and Future Horizons*. [with Stephen Stich]
'Racial Cognition and Normative Racial Theory,' to appear in *The Handbook of Moral Psychology*. [with Edouard Machery and Ron Mallon]
'Moral Judgment,' to appear in the *Routledge Companion to the Philosophy of Psychology*. [with Jenny Nado and Stephen Stich]