This article appeared in *Cognition* (2002) 84, 221-236.

Norms with Feeling: Towards a Psychological Account of Moral Judgment

Shaun Nichols[a]

[a]*Department of Philosophy, College of Charleston, Charleston, SC 29424*

Abstract

There is a large tradition of work in moral psychology that explores the capacity for moral judgment by focusing on the basic capacity to distinguish moral violations (e.g., hitting another person) from conventional violations (e.g., playing with your food). However, only recently have there been attempts to characterize the cognitive mechanisms underlying moral judgment (e.g., Blair 1995, Goldman 1993). Recent evidence indicates that affect plays a crucial role in mediating the capacity to draw the moral/conventional distinguish. However, the prevailing account of the role of affect in moral judgment is problematic. This paper argues that the capacity to draw the moral/conventional distinction depends on both a body of information about which actions are prohibited (a Normative Theory) and an affective mechanism. This account leads to the prediction that other normative prohibitions that are connected to an affective mechanism might be treated as nonconventional. An experiment is presented that indicates that "disgust" violations (e.g., spitting at the table), are distinguished from conventional violations along the same dimensions as moral violations.

Keywords: Moral/conventional distinction, moral judgment, norms, moral psychology, disgust

---

1. Introduction

Many of the deepest issues concerning the nature of morality would be illuminated if we had an adequate account of the nature of moral judgment. So it is scarcely surprising that psychologists and philosophers have invested enormous effort in trying to produce an account of moral judgment. The exploration of moral judgment in psychology stretches back for a century, through Kohlberg and Piaget. The philosophical lineage is much longer and enjoys an even more distinguished cast, including Kant, Hume, and Aristotle. Despite this rich history of research on moral judgment, only recently have there been sustained attempts to characterize the cognitive mechanisms underlying moral judgment (e.g., Blair, 1995; Goldman, 1993).

The basic capacity for moral judgment has perhaps been most directly approached empirically by using the moral/conventional task, which explores subjects' ability to distinguish moral from conventional transgressions. The spate of empirical literature on this issue began with the work of Elliot Turiel (e.g., Turiel, 1983), and the research program generated by Turiel's work indicates that people distinguish moral violations (e.g., pulling hair) from conventional violations (e.g., drinking soup out of a bowl). What is striking about this literature is that, from a young age, children distinguish cases of moral violations from cases of conventional violations on a number of dimensions. For instance, children tend to think that moral transgressions are generally less permissible and more serious than conventional transgressions. And the explanations for why moral transgressions are wrong are given in terms of fairness and harm to victims, whereas the explanations for why conventional transgressions are wrong are given in terms of social acceptability. Further, conventional rules, unlike moral rules, are viewed as dependent on authority. For instance, if at another school the teacher has no rule against chewing gum, children will judge that it's not wrong to chew gum at that school; but even if the teacher at another school has no rule against hitting, children claim that it's still

wrong to hit at the school. These findings on the moral/conventional distinction have turned out to be quite robust. They have been replicated numerous times using a wide variety of stimuli (see Smetana, 1993 and Tisak 1995 for reviews). Thus, the capacity for drawing the moral/conventional distinction plausibly indicates a basic capacity for moral judgment.

Recent evidence from Blair (1995) suggests that affect plays a crucial role in this basic capacity for moral judgment. In what follows, I will consider Blair's account of moral judgment and offer an alternative proposal. I'll argue that the capacity to draw the moral/conventional distinction depends on both a body of information about which actions are prohibited (a Normative Theory) and an affective mechanism that confers a special status on the norms. In support of this view, I present empirical evidence that prohibitions against disgusting actions (e.g., spitting at the table) are distinguished from conventional violations along the same dimensions as moral violations.

2. Blair's VIM-account

Armed with a dazzling series of experiments, Blair has developed the most detailed cognitive account of the role of affect in moral judgment. Blair maintains that the capacity to draw the moral/conventional distinction derives from the activation of a Violence Inhibition Mechanism (VIM). The idea for VIM comes from Konrad Lorenz' (1966) suggestion that social animals like canines have evolved mechanisms to inhibit intra-species aggression. When a conspecific displays submission cues, the attacker stops. Blair suggests that there's something analogous in our cognitive systems, the VIM, and that this mechanism is the basis for our capacity to distinguish moral from conventional violations.

Unfortunately, it's not entirely clear how VIM is supposed to produce the

moral/conventional distinction, but we do get a broad outline from Blair (1995). It's useful to divide Blair's theory into two parts. The first part of the theory proposes that VIM generates a sense of aversion. VIM is activated by distress cues, but VIM-activation initially simply produces a withdrawal response. This VIM-activation becomes aversive through "meaning analysis": "the withdrawal response following the activation of VIM is experienced, through meaning analysis, as aversive" (1995, p. 7). There are important questions about what the meaning analysis comes to, but Blair (1995) does not elaborate this part of his theory.[1] Nonetheless, the important point for our purposes is that the aversive feeling depends on both VIM and meaning analysis. The second part of Blair's theory is that it is this feeling of aversiveness that generates the responses to the moral items on the moral/conventional task. According to Blair, VIM (plus meaning analysis) produces an aversive experience and it is "this sense of aversion to the moral transgression" that results in the act being "judged as bad" (1995, p. 7). On Blair's account, then, the process seems to go as follows. The VIM is triggered by distress cues or by associations to distress cues; this VIM activation is experienced as aversive through meaning analysis; and events that are experienced as aversive in this way are treated as nonconventional transgressions in the moral/conventional task (Blair, 1995, p. 7; Blair, 1993, pp. 83, 88).

[1] Blair (1993) adverts to Mandler's definition of meaning analysis as "the activation and accessibility of those schematic representations that best fit the available evidence" (Mandler 1984, 126). The meaning analysis seems to supply the agent with certain interpretations about events, and these interpretations then play a crucial role in the generation of conscious emotional states (Blair 1993, pp. 61, 83; Mandler 1984, pp. 46, 126). In the case of VIM, apparently the interpretation of VIM-activation leads to the experience of aversion.

Blair's primary evidence for his theory comes from a series of studies on psychopaths. He presented the moral/conventional task to psychopaths in British prisons (Blair, 1995; see also Blair 1997). Since the pool of psychopaths was drawn from a prison population, Blair used non-psychopathic prison inmates as a control. Blair found that control criminals made a significant moral/conventional distinction on permissibility, seriousness, and authority dependence; psychopaths, on the other hand, didn't make a significant moral/conventional distinction on any of these dimensions. Further, although the control criminals tended to appeal to the victim's welfare to explain why the moral transgressions were wrong, psychopaths tended to give conventional-type justifications for both the moral and the conventional transgressions. Apparently, then, the capacity for moral judgment is compromised in psychopathy. Blair and colleagues also found another important difference between psychopaths and control criminals. Non-psychopathic criminals show high physiological response to cues of distress in others. By contrast, psychopaths show significantly lower physiological response to distress cues (Blair et al., 1997). Blair interprets this as evidence that psychopaths have a defective VIM, and thus that the evidence supports his account of moral judgment.

One important feature of Blair's account is that it proposes that VIM is independent of any capacity for understanding other minds, or 'mindreading'. Hence, on Blair's account, it is possible for the capacity to draw the moral/conventional distinction to be dissociated from the capacity for mindreading. Blair tries to support this claim by appealing to his finding that autistic children were able to make the moral/conventional distinction, despite their difficulties with mindreading (Blair, 1996). He suggests that this evidence shows that the capacity for mindreading or 'mentalizing' is entirely dissociated from the capacity to draw the moral/conventional distinction: "Children with autism have been demonstrated to be incapable of

'mentalizing' (e.g., Baron-Cohen, Leslie & Frith, 1985)" and so, they are incapable of "forming a representation of the mental state of the other" (Blair, 1995, p. 22). He maintains that his theory explains how autistic children can make the moral/conventional distinction even though they can't mentalize: "While children with autism may not be able to represent a mental state of another's distress, this distress, as a visual or aural cue, will activate their VIM" (Blair, 1995, p. 22).

Blair's argument for a dissociation between the capacity for mindreading and the capacity for the moral/conventional distinction is unconvincing. Claiming that autistic children can't "mentalize" or that they can't represent the mental states of others overstates their deficit. There is reason to think that autistic children can represent some mental states. Autistic children are capable of attributing simple desires and emotions (e.g., Tan & Harris, 1991; Yirmiya et al., 1992). They understand that people can have different desires and "that someone who gets what he wants will feel happy, and someone else who does not get what he wants will feel sad" (Baron-Cohen, 1995, p, 63). Furthermore, studies of spontaneous language use in autistic children indicate that these children use the term 'want' and 'hurt' appropriately (Tager-Flusberg, 1993). Thus there is good reason to think that the capacity for attributing desires is largely intact in autistic children (see also Nichols & Stich forthcoming). As a result, the fact that these children can distinguish moral from conventional violations does not provide evidence that the capacity for making this distinction is entirely independent from the capacity for mindreading. Of course, one of the major themes in recent work on mindreading is that it is important to distinguish among different aspects of mindreading. For instance, the capacity for attributing beliefs might depend on different mechanisms from the capacity for attributing desires and emotions (Nichols & Stich forthcoming). So Blair's evidence might be taken to

support the more restricted claim that *some* aspects of mindreading are dissociable from the capacity for drawing the moral conventional distinction. For instance, the evidence on autism might be taken to support the view that the capacity to attribute false beliefs is dissociable from the capacity to draw the moral/conventional distinction. However, Blair's evidence does not support the stronger claim that the capacity for drawing the moral/conventional is dissociable from all mindreading capacities. In particular, the evidence on autism does not support the claim that the capacity for drawing the moral/conventional distinction is dissociable from the capacity to represent the mental states of another's distress. For there is good reason to think that this mindreading capacity is intact in autism.

So, Blair's evidence does not confirm his hypothesis that mindreading is unnecessary for drawing the moral/conventional distinction. Moreover, there are serious shortcomings in Blair's account itself. Perhaps the easiest way to illustrate the shortcomings is by exploiting the important distinction between judging something *bad* and judging something *wrong*. Many occurrences that are regarded as bad are not regarded as wrong. Toothaches, for instance, are bad, but they aren't wrong. The moral/conventional task gets its interest primarily because it gives us a glimpse into judgments of *wrong*. This is reflected by the fact that the items in the moral/conventional task are explicitly *transgressions*, and the very first criterion category is *permissibility*. As we'll see, the problem with Blair's account is that, while the proposal might provide an account of judging something bad (in a certain sense), it does not provide an account of judging something wrong.

If the first part of Blair's theory is right, VIM (plus meaning analysis) produces a distinctive aversive response. As with toothaches, we might regard the stimuli that prompt this aversive response as 'bad'. Furthermore, it might be important to treat stimuli that produce

VIM-based aversion as 'bad' in a distinctive way. Now, what is the class of stimuli that are bad in this sense? Well, anything that reliably produces VIM activation. Distress cues will be at the core of this stimulus class (Blair, 1995, 1999). The class of stimuli that will be accordingly aversive will include distress cues from victims of natural disasters and accidents and even superficial distress cues like paintings and drawings. Thus, the class of stimuli that VIM (plus meaning analysis) will lead us to regard as "bad" includes natural disaster victims, accident victims, and superficial distress cues. But it is quite implausible that these things are *wrong.* Natural disasters are, of course, bad. But, barring theological digressions, natural disasters aren't regarded as *wrong*. Indeed, this is clear from the first criterion category in the moral/conventional task – it doesn't even make sense to say that natural disasters are impermissible. Similarly, if a child falls down, skins her knee, and begins to cry, this will produce aversive response in witnesses through VIM. Yet the child's falling down doesn't count as a moral transgression. It's instructive here to consider a probe that is sometimes used to distinguish between transgressions and non-transgressions. Subjects asked whether *punishment* is appropriate for certain events tend to say that punishment is appropriate for both conventional and moral transgressions, but not for *nontransgressions* (cf. Davidson et al. 1983, Zelazo et al. 1996). Although a child crying after a fall will be "bad" in the VIM sense, it would be rather sadistic to suggest that the child should be punished. This is plausibly because scraping one's knee and crying is not considered *wrong.* As a final example, consider again the fact that superficial distress cues produce aversive response through VIM (plus meaning analysis). This aversive response can be generated whether or not one believes that the other person is in distress. Indeed, this is crucial for Blair's view on autism and moral judgment. According to Blair, even though autistic children cannot represent distress, they have an intact VIM, and this is

the basis for their capacity for moral judgment. Artificial distress cues might thus be judged as bad in the VIM sense, but producing such cues (e.g., by creating or playing a tape of simulated crying) would presumably not be judged as nonconventionally wrong. That is, it's unlikely that producing artificial distress cues would be judged to be significantly less permissible, more serious and less authority contingent than standard conventional transgressions (e.g., wearing pajamas to class). Indeed, like the cases of natural disasters and accidental injury, producing artificial distress cues is typically not regarded as a transgression at all.

So, while Blair's theory might provide an account of how people come to judge things as *bad* in a certain sense, it does not provide an adequate account of moral judgments of *wrong* on the moral/conventional task. Of course, Blair's theory might be developed further to try to exclude all of the problematic cases, but as it stands, the theory has no motivated explanation for why these "bad" stimuli aren't regarded as wrong.

*3. Moral judgment depends on an **Affect-Backed Normative Theory***

The central problem for Blair's theory, I've argued, is that it does not provide an adequate account of judgments of wrong. Perhaps the most plausible way to remedy this problem is to maintain that there is a body of information specifying which acts are wrong, i.e., which acts are transgressions. On this proposal, in typical moral scenarios presented in the moral/conventional task, people's judgments are guided by a body of information, a 'Normative Theory' prohibiting behavior that harms others.[2] However, the Normative Theory surely does not consist of a single simple rule. For instance, at least among adults, the Normative Theory

[2] The most familiar instances of such harm-norms prohibit psychological harms like pain and suffering.

allows that it is sometimes acceptable to harm a child for her long term benefit.

Part of what makes the appeal to a Normative Theory plausible is that it's widely agreed that all of the populations studied in these tasks have information about normative prohibitions – for all the populations are fluent with the *conventional transgressions*, and the prevailing explanation of this is that subjects have knowledge of the conventional rules. Furthermore, despite their problems on the moral/conventional task, psychopaths evidently have a body of knowledge about the rules prohibiting harming others. The suggestion I'm promoting is that psychopaths aren't unusual in having a Normative Theory proscribing harming others. Everyone who is competent on the moral/conventional task has a Normative Theory proscribing harming others.[3]

The Normative Theory that prohibits harming others, on the current proposal, does depend on some capacity for mindreading. For it requires some mindreading abilities to properly categorize harm and to recognize the distinction between genuine and superficial distress cues. Nonetheless, the requisite mindreading abilities here are plausibly quite minimal. As a result, the evidence on autism fits comfortably with the present proposal. For, as noted earlier, despite their deficits in some aspects of mindreading, autistic children are capable of attributing wanting and hurting to others.[4] Of course, this points to an important difference between the present proposal

[3] Although this goes well beyond the scope of this paper, there is a wide range of unanswered questions about the Normative Theory, including the following: Does the Normative Theory have innate constraints? Does it have universal components? Is the information encapsulated?

[4] This does not exclude the possibility that there might be important differences between the moral judgments of children with autism and those of other children. On the account proposed here, one of the central features of moral development is that the Normative Theory becomes

and the available options for Blair's account. For Blair's account explicitly forswears any use of mindreading; as a result, the cases of superficial distress cues cannot be excluded by knowledge that there is no genuine suffering.

An adequate account of moral judgment must also explain Blair's data on psychopaths. Blair finds that psychopaths have a deficit both in moral judgment and in their affective response to others' suffering. Although Blair characterizes the affective deficit as a deficit to VIM, VIM is linked to Lorenz' evolutionary account, which is regarded with considerable suspicion in the contemporary literature (e.g. de Waal 1996). As a result, I'm inclined to adopt a descriptive characterization of the affective system that is neutral about evolutionary function. In the literature on prosocial behavior, it is common to distinguish between two kinds of emotional response to another's suffering: personal distress and concern (see, e.g., Batson, 1991; Eisenberg et al., 1989). Both of these kinds of responses seem to emerge early in development, and it's likely that the mechanisms underlying these responses are defective in psychopathy (see, e.g., Nichols, 2001). So, the affective mechanism important for moral judgment might be a mechanism for concern or personal distress rather than VIM. For current purposes, the precise characterization of the affective mechanism is not really crucial. The important claim is simply that some affective mechanism that is responsive to others' suffering is plausibly implicated in moral judgment.

---

increasingly sophisticated. Presumably this increasing sophistication will sometimes draw on increasingly sophisticated mindreading abilities. For instance, it is part of the Normative Theory of older children (and adults) that lying is prohibited. But since an understanding of lying depends on fairly sophisticated mindreading capacities, this prohibition may be absent from the autistic child's Normative Theory.

Thus, I suggest that moral judgment depends on two mechanisms, a Normative Theory prohibiting harming others, and some affective mechanism that is activated by suffering in others. On this account, then, the system underlying moral judgment is what we might call an Affect-Backed Normative Theory in which the Normative Theory prohibits actions of a certain type, and actions of that type generate strong affective response.[5] There is reason to think that the two mechanisms underlying moral judgment are at least partly dissociable. Children exhibit both personal distress and concern well before the second birthday (e.g., Simner, 1971; Zahn-Waxler et al., 1992). But 1-year olds presumably do not make moral judgments, and this can be attributed to the fact that they have not yet developed an understanding of the Normative Theory that will guide their moral judgments in the coming years. Psychopaths, on the other hand, seem to have a dissociation in the other direction. They show a deficit in affective response to suffering, and this seems to compromise their ability to respond normally on the moral/conventional task. But psychopaths apparently have a largely intact knowledge of the rules prohibiting harming others.

[5] I don't mean to suggest that conventional transgressions carry no affective force. People might find it generally upsetting when rules of any sort are broken. But of course, since this applies to all rules, it doesn't distinguish conventional normative judgment from nonconventional normative judgment. My claim might be somewhat more carefully cast, then, as the claim that moral violations implicate an affective component that goes beyond whatever affect might attend all transgressions. This still leaves open important questions. As an anonymous referee suggested, transgressions might come to be treated as nonconventional because of the intensity of the affect or, alternatively, because of the *kind* of affect.

4. The disgusting/conventional distinction: some empirical results

I've claimed that the nonconventional responses to the moral questions derive from two factors, a Normative Theory prohibiting harm and an affective system that is sensitive to harm in others. On the account of moral judgment that I've proposed, the moral/conventional task really taps a distinction between a set of norms that are backed by an affective system and a set of norms that are not backed by an affective system. On this theory, affect-backed normative claims will be treated differently than affect-neutral normative claims. Thus, the account predicts that transgressions of other (non-harm-based) rules that are backed by affective systems should also be treated as nonconventional. As a result, if we find that other affect-backed norms are also distinguished from conventional norms along the dimensions of permissibility, seriousness, authority contingency and justification type, then this will provide an independent source of evidence for the account of moral judgment that I've proposed. On the VIM account, by contrast, treating transgressions as nonconventional depends on the VIM, so this account does not predict that transgressions implicating other emotional responses will be treated as nonconventional.

To test the prediction we need to exploit a body of (non-harm-based) rules that are backed by an affective system. Over the course of two experiments, I explored the extent to which normative violations involving disgust would be distinguished from affectively neutral normative violations.

Experiment 1

For this experiment, subjects were given a set of transgression scenarios, each of which was followed by questions about permissibility, seriousness, authority contingency and

justification. Two of the scenarios were moral transgressions, two were neutral conventional transgressions, and two were disgusting transgressions. The hypothesis was that subjects would distinguish disgusting transgressions from neutral conventional transgressions on all the criterion judgments (i.e., permissibility, seriousness, authority contingence) and that subjects would tend to give different kinds of justifications for the two classes of violations. More specifically, the directional hypotheses were that subjects would judge disgust-violations to be less permissible, more serious and less authority contingent than the neutral-violations.

Method

Participants

19 students from an introductory philosophy course at the College of Charleston participated in this study.

Materials

The moral and conventional scenarios used were taken from the literature. The two moral stories involved one child hitting another child and one child pulling another child's hair. The two conventional stories involved a child wearing pajamas to school and an adult drinking tomato soup out of the bowl at a dinner party. In addition to the standard moral and conventional stories, subjects were given two disgust stories. In one of these stories, a child puts her finger in her nose in class. In the other story, a person at a dinner party spits in his water glass before drinking it. The order of the disgusting and conventional stories was counterbalanced.

Procedure

Subjects were given questionnaires in a classroom. The questionnaires contained 6 scenarios: two moral, two conventional, and two disgusting. In each case, after the transgression is described, the subject is asked 4 questions. For instance, in one of the disgust-scenarios, subjects are presented the following scenario and questions:

> Bill is sitting at a dinner party and he snorts loudly and then spits into his water before drinking it.
>
> 1. Was it O.K. for Bill to spit in his water?
>
> If it's not O.K. for Bill to do that, then:
>
> 2. On a scale of one to ten, how bad was it for Bill to spit in his water?
>
> 3. Why was it bad for Bill to spit in his water?
>
> 4. Now what if, before Bill went to the party, the hosts had said, "At our dinner table, anyone can spit in their food or drink." Would it be O.K. for Bill to spit in his water if the hosts say he can?

Results

Scoring

The scoring procedure follows that of the previous studies in the literature (Smetana & Braeges, 1990; Blair, 1995). Questions 1 and 4 were scored binomially, with each No answer being given a score of 1, so the cumulative score for each domain could range from 0 to 2. Question 2 was coded by the value (between 1 and 10) given to the seriousness of the transgression. Question 3 was coded according to the justification categories in table 1. Two independent coders scored the justifications, and inter-rater reliability was high (82% for all

items; 86% for disgusting and conventional items).

Table 1: Description of Justification Categories:

Other's welfare	Any reference to victim's welfare (e.g., "it will hurt her"; "it's not fair")
Rule	Any reference to rules, even if implicit (e.g., "it's not socially acceptable")
Disorder	Any reference to disorder caused by the behavior (e.g., "it will distract others")
Rudeness	Any reference to the rudeness of the behavior (e.g., "it's bad manners")
Health	Any reference to health risks involved with the behavior (e.g., "bad hygeine")
Disgust	Any reference to the disgustingness of the behavior (e.g., "it's gross")
Other	Any other response

Analysis

The hypothesis that disgusting transgressions would be distinguished from the neutral conventional transgressions was confirmed. The disgusting violations were distinguished from the conventional violations on all the criterion judgments. The disgusting violations were regarded as less permissible (McNemar's test, N=19, $p < .025$), more serious ($t(4) = 2.954$, $p < .05$) and less authority contingent (McNemar's test, N=13, $p < .025$) than the neutral violations (all tests two-tailed).

Although the results here reach significance, there are a couple of shortcomings with this analysis. First, for the seriousness question, the N is quite small since if the subject said that one of the actions was permissible, then that subject typically did not answer the subsequent seriousness question. As a result, those subjects were not used in the cumulative analysis, and

there were only 5 remaining subjects who did answer the seriousness question for all conventional and disgusting items. The other problem with the findings is that the justification question turned up a potentially important difference in response between the disgust-violations themselves. In the finger-nose case, only three subjects appealed to the disgustingness of the action to explain why it was wrong. A greater number appealed to health considerations (e.g., germs, unsanitary behavior) to explain why the action was wrong. Since this raises the possibility that responses to this probe are not mediated by disgust, analyses were also performed on the spitting case alone, compared with the soup case (which is most closely matched to the spitting case). The results were still quite significant. Spitting in the water glass before drinking was judged as less permissible than drinking the soup (McNemar's test, N=19, $p < .05$); it was regarded as more seriously wrong than drinking the soup ($t(10)=5.328$, $p < .01$); and the transgression was regarded as less dependent on authority than drinking the soup (McNemar's test, N=15, $p < .025$). Furthermore, in the spitting case, over 60% of the subjects explained why the action was wrong by appealing to disgust (e.g., "because that's gross!"). None of the subjects gave this sort of explanation for why it is wrong to drink the soup. Rather, the subjects offered conventional justifications for why drinking the soup was wrong, either adverting to rudeness (e.g., "It's bad manners") or to rules (e.g., "You aren't supposed to do that at social functions"). Although there were significant differences between the spitting case and the soup case on all the criterion items, there were no significant differences between the spitting case and the pulling hair case on permissibility (N=19, $p = .5$, n.s.), seriousness (t (16) = -1.127, $p = .276$, n.s.) or authority contingency (N=16, $p = .25$, n.s.). Of course, the justifications for why the actions were prohibited were quite different. Subjects typically offered welfare-based explanations for why pulling hair is wrong, and, as we've seen, they offered disgust-based

explanations for why spitting in the glass before drinking is wrong.

Experiment 2

The preceding experiment shows that disgusting violations, which seem clearly to be affectively charged, are treated as quite distinct from neutral conventional violations. What the experiment does not address, however, is whether diminished responsiveness to disgust will have an effect on a subject's tendency to treat disgusting violations as impermissible, serious or authority independent. Hence, for experiment 2, I wanted to see whether such differences might be revealed. In this experiment, two groups of subjects, subjects with high disgust sensitivity and subjects with low disgust sensitivity, were compared on their responses to the permissibility, seriousness and authority contingency of a disgusting violation. The hypothesis was that the high disgust subjects would be more likely than low disgust subjects to judge disgust-violations as impermissible, very serious and not contingent on authority.

Method

Participants:

24 undergraduates from an introductory philosophy class at the College of Charleston participated in this study. Subjects were divided into high- and low-disgust sensitivity groups on the basis of their scores on the Disgust Scale (Haidt et al., 1994). A median split on subjects' scores yielded two groups of 12 subjects.

Materials:

In constructing the probes for Experiment 1, I anticipated that the disgust probes would be so disgusting that there would be little variation in the response. This turned out to be right, especially for the "spitting in the glass" probe. The vast majority of subjects regarded this behavior as impermissible, very serious and not authority contingent. As a result, this scenario leaves little room to find variation between subjects, so I prepared different, slightly less disgusting, probes for this experiment. For instance, the spitting in the glass probe was replaced with the following:

> Michael is sitting at a dinner party and he picks up a paper napkin, snorts, and spits into the napkin.

The subjects were subsequently asked to judge the permissibility, seriousness, justification, and authority contingence of this action, as in experiment 1. Subjects then filled out the disgust scale questionnaire (Haidt et al., 1994).

Procedure:

The procedure was the same as in experiment 1.

Results

Scoring:

The scoring for the Disgust Scale was the standard scoring method described in Haidt et al, 1994. Otherwise, the scoring was the same as in experiment 1.

Analysis:

There were two disgust-violation questions. However, one of these questions, which

involved a girl looking into a tissue after blowing her nose, produced little variation in the subjects. The vast majority of subjects thought that the action was permissible and only 3 of the 24 subjects claimed that the action was not authority contingent. So this left little opportunity to explore individual differences. Fortunately, for the other disgust-violation question, spitting into the napkin, there was considerable variation. Fourteen subjects maintained that the action was not authority contingent and ten maintained that it was. So the analyses on the relation between disgust-sensitivity and the criterion judgments were performed on this probe. There was no statistically significant difference between low & high disgust subjects on the permissibility question (χ^2_{corr}(N=24, df=1) = .300, p=.146, n.s.). However, there were significant group differences on the other two criterion judgments. The low disgust subjects were more likely than the high disgust subjects to treat the transgression as authority contingent, χ^2_{corr}(N=24, df=1) = 4.286, p< .05. (See table 2.)

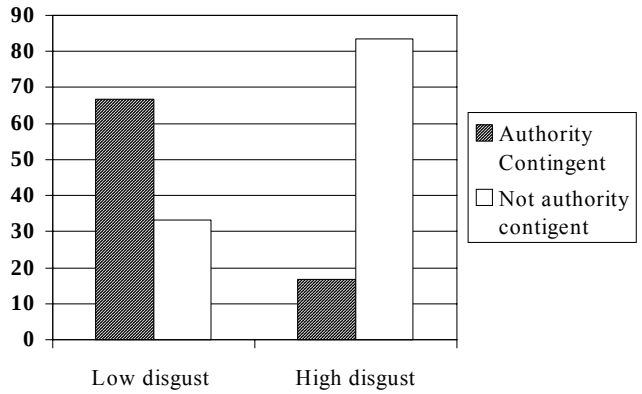

Table 2: Judgments of authority contingence in low and high disgust subjects

Further, low disgust subjects judged the transgression as significantly less serious than high disgust subjects, $t(19) = -2.699$, $p < .025$. (See table 3.)

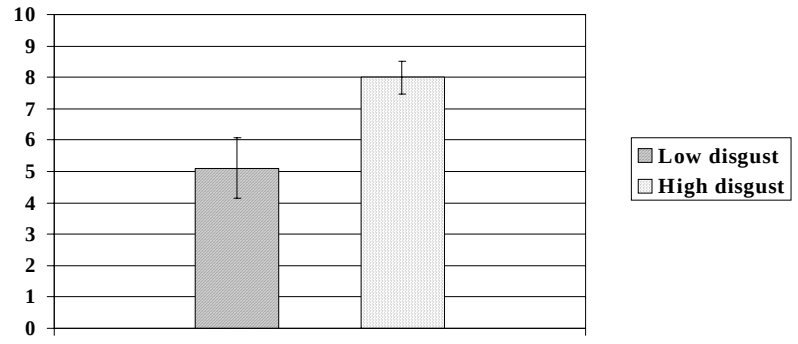

Table 3: Judgments of seriousness in low and high disgust subjects

In addition to analyzing group differences between low-disgust and high-disgust groups, I also conducted correlational analyses, and the effects were comparable. There was a significant correlation between disgust scale scores and answering that the transgression is not authority contingent ($r_{pb} = +.47$, N=24, p<.05). There was also a significant correlation between disgust scale scores and the seriousness of the transgression ($r = +.54$, N=21, p<.05). The correlation between disgust scale scores and authority contingence remained significant after partialing out gender using the Pearson R partial correlation ($r = +.42$, p<.05). The correlation between disgust

scale scores and seriousness also remained significant after partialing out gender (r = +.63, p<.01).

5. Norms with feeling: The disgusting and the immoral

The findings of Experiment 1 confirmed the prediction that disgust-backed transgressions would be distinguished from affectively neutral transgressions on the classic moral/conventional dimensions. Transgressions that are disgust-backed are judged to be less permissible, more serious, less contingent on authority, and are more likely to elicit nonconventional justifications than affectively neutral conventional transgressions. The findings of experiment 2 indicate that the affective response does play a critical role in prompting individuals to treat disgusting violations as nonconventional. For experiment 2 revealed that low-disgust subjects were more likely than high-disgust subjects to judge the disgusting violation as contingent on authority and less serious. This indicates that responses to the criterion dimensions are somehow mediated by affective response. Although the results thus fit the predictions generated by the Affect-Backed Theory account, this evidence does not fit comfortably with the VIM account. For the disgusting actions are not distress cues, and as a result will not activate VIM. Nonetheless disgusting transgressions are treated as nonconventionally wrong. As a result, the evidence indicates that VIM is not necessary for treating transgressions as nonconventionally wrong.[6]

[6] Blair (personal communication) has pointed out that my proposal generates the prediction that psychopaths will distinguish disgust transgressions from conventional transgressions. It will be interesting to see whether this is borne out by future experiments. Of course, the prediction that psychopaths will distinguish disgusting transgressions from conventional transgressions depends on the assumption that psychopaths have normal levels of disgust sensitivity. For experiment 2

It's worth saying a bit more clearly how the findings help to confirm the Affect-Backed Theory framework suggested above for moral judgment. Since the experiments indicate that the disgust system provokes nonconventional responses to questions about permissibility, seriousness, authority contingency and justification, we have evidence that nonconventional responses to these questions *can be* induced by affective response. There is independent reason to think that suffering in others inspires considerable affective response, and that this kind of affective response to others' distress emerges quite early (see e.g., Nichols, 2001; Zahn-Waxler et al, 1992). As a result, we have a couple of important pieces in place to corroborate the Affect-Backed Theory account of moral judgment. Harm-scenarios generate affective response, and affective response can provoke nonconventional answers to the standard moral/conventional questions. So it's reasonable to suppose that the affective response to harm-scenarios does play a crucial role in leading subjects to judge that hitting others and pulling hair is impermissible, very serious, and not authority contingent. More broadly, it is plausible that the norms prohibiting disgusting behavior and the norms prohibiting harmful behavior are part of an important class of norms, "norms with feeling". Violations of norms with feeling are judged as less permissible, more serious, and less dependent on authority than conventional normative violations. In addition, the level of affective response has a significant effect on the extent to which subjects distinguish norms with feeling from norms without.[7]

---

indicates that subjects with low levels of disgust sensitivity are more likely to regard disgusting violations as authority contingent and less serious.

[7] Norms that are connected to affective response in this way might also accrue an advantage in cultural evolution – i.e., they may be more likely to survive than norms that are not connected to affective response (Nichols 2002).

Although affect thus seems to play a crucial role in generating nonconventional judgment, such judgment cannot be wholly explained by appealing to affect, even in the case of "disgust" transgressions. Rather, just as there are norms against harming others, there are norms against disgusting behavior. And, as in the case of harm norms, there isn't a single simple rule that prohibits disgusting behaviors. For clearly some disgusting behaviors, e.g., unintentional vomiting, are not prohibited. There are even intentional actions that are disgusting but not prohibited, e.g., some parlor tricks (I leave it to the reader to recall or construct his own examples). Thus, there seems to be a body of information, a normative theory, proscribing a class of disgusting behavior. Furthermore, it seems that the normative theory is at least partially independent from the affective system. For even though low disgust subjects were more likely to say that the disgusting action was not wrong if an authority said it wasn't, most of these subjects also maintained that the action was normatively prohibited.

On the general picture that I've suggested, then, two quite different mechanisms are implicated in nonconventional normative judgment: a normative theory and an affective system. This proposal leaves open a wide range of possibilities about how these mechanisms work together to produce nonconventional normative judgment. The evidence suggests that affective mechanisms can play a crucial role in generating nonconventional judgments. But *what* role does affect play? In particular, what role does affect play in leading subjects to judge disgusting violations and moral violations as (i) very serious, (ii) not contingent on authority, and (iii) possessing non-conventional justification? There isn't enough evidence available to answer these questions with any confidence. But one explanation for why disgusting transgressions are judged as less permissible and more serious than affectively neutral transgressions is that

disgusting transgressions carry both the aversiveness of being transgressive and an additional aversive component – they provoke disgust. That is, in addition to violating a rule, disgusting transgressions activate an affective mechanism, which makes them more offensive than transgressions that merely violate a rule. This also suggests how disgusting violations might come to be regarded as not authority contingent and as having extra-conventional justification. In the case of conventional violations, when we imagine that the rules are suspended, that suffices to undermine any basis for judging the action as an offense. By contrast, when one imagines that certain disgust rules are suspended, as in the case of the host proclaiming that it is okay to spit in one's water glass, imagining the activity still provokes the disgust response and so this activity continues to be regarded as an offense. Indeed, on this proposal, precisely what makes disgust-transgressions especially serious persists even when the rules are suspended. What makes them especially serious is that these violations are *disgusting*.

This is, of course, a rather crude processing model for what is surely a rich and complex phenomenon. One interesting possibility is that there are important ontogenetic factors in fixing the cognitive mechanisms that subserve judgments surrounding disgusting violations and harmful violations. There might be a critical period in development during which affective mechanisms combine with information about normative prohibitions to form a kind of nonconventional normative theory. For instance, in the case of disgusting violations, the child is provided with (whether by instruction or by innate endowment) a body of rules against disgusting behavior. The disgust mechanism might then shape this body of rules into a nonconventional normative theory, which is importantly different from the other, conventional normative theories that the child develops, like the normative theory about table manners. More generally, it might be that when normative prohibitions are paired with affective response, as is

the case with disgust prohibitions, the affect provides a kind of reinforcement for the prohibitions that instills a deeper repugnance for actions that transgress these norms, and this might infuse the norms with a nonconventional status.

The available evidence does not remotely decide whether this is the right sort of account for how affect produces nonconventional responses to disgusting violations, much less for moral violations. The point I wish to emphasize is simply that even if nonconventional normative judgment does involve both a normative theory and an affective mechanism, it remains to be seen exactly how those different mechanisms conspire to enable the distinctive kinds of judgments subjects make about disgusting violations and moral violations.

The idea that affect is essential to moral judgment has extremely distinguished intellectual origins, tracing back to the philosophical musings of Hume and Shaftesbury. The ensuing centuries have seen philosophers feverishly debate whether moral judgment really does depend on the emotions. This philosophical debate has not, to say the least, produced consensus. Only now does it seem that cognitive science is poised to build an empirical case that would vindicate Hume's speculation.

Acknowledgements

I am grateful to Justin D'Arms, Owen Flanagan, Trisha Folds-Bennett, Justin Halberda, James Hittner, Ron Mallon, Kim May, Elizabeth Meny, and Steve Stich for discussion and comments on earlier versions of this paper. I would also like to thank three anonymous referees for their comments. And I owe special thanks to James Blair for providing extensive feedback on previous drafts. Versions of this paper were presented at Washington University, the

University of Louisiana at Lafayette, Hampshire College, and at the Society for Philosophy and Psychology. I am grateful to the audiences on those occasions for their helpful responses.

References

Baron-Cohen, S. (1995). *Mindblindness,* Cambridge, MA: MIT Press, Bradford Books.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a "Theory Of Mind"? *Cognition*, 21, 37 46.

Batson, C. (1991). *The altruism question.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Blair, R. (1993). *The development of morality*. Unpublished Ph.D. thesis, University of London.

Blair, R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, 57, 1-29.

Blair, R. (1996). Brief report: morality in the autistic child. *Journal of Autism and Developmental Disorders*, 26, 571-579.

Blair, R. (1997). Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences*, 26, 731-739.

Blair, R. (1999). Psychophysiological responsiveness to the distress of others in children with autism. *Personality & Individual Differences*, 26, 477-485.

Blair, R., Jones, L., Clark, F., Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, 34, 192-198.

Davidson, P. Turiel, E., and Black, A. (1983). The effect of stimulus familiarity on the use of criteria and justifications in children's social reasoning, *British Journal of Developmental Psychology*, 1, 49-65.

De Waal, F. (1996). *Good natured,* Harvard University Press.

Eisenberg, N., Fabes, R., Miller, P., Fultz, J., Shell, R., Mathy, R. and Reno, R. (1989). Relation of sympathy and personal distress to prosocial behavior: a multimethod study. *Journal of Personality and Social Psychology*, 57, 55-66.

Goldman, A. (1993). Ethics and cognitive science. *Ethics* 103, pp. 337-360.

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences*, 16, 701-713.

Lorenz, K. (1966). *On aggression,* New York: Harcourt, Brace, Jovanovich.

Mandler, G. (1984). *Mind and body.* New York: Norton & co.

Nichols, S. (2001). Mindreading and the cognitive architecture underlying altruistic motivation. *Mind & Language,* 16, 425-455.

Nichols, S. (2002). On the genealogy of norms: A case for the role of emotion in cultural evolution. *Philosophy of Science,* 69.

Nichols, S. and Stich, S. forthcoming. *Mindreading*. Oxford: Oxford University Press.

Simner, M. (1971). Newborn's response to the cry of another infant. *Developmental Psychology 5*, 136-150.

Smetana, J. (1993). Understanding of social rules. In M. Bennett (ed.) *The development of social cognition : the child as psychologist*. New York: Guilford Press, 111-141.

Smetana, J. & Braeges, J. (1990). The development of toddlers' moral and conventional judgements. *Merrill-Palmer Quarterly*, 36, 329-346.

Tager-Flusberg, H. (1993). What language reveals about the understanding of minds in children with autism. In S. Baron-Cohen, H. Tager-Flusberg & Donald Cohen (eds.) *understanding other minds: perspectives from autism*, 138-157.

Tan, J., and Harris, P. (1991). Autistic children understand seeing and wanting. *Development and Psychopathology*, 3, 163-174.

Tisak, M. (1995). Domains of social reasoning and beyond. In R. Vasta (ed.), *Annals of child development*, Vol. 11 (pp. 95-130). London: Jessica Kingsley.

Turiel, E. (1983). *The development of social knowledge: morality and convention,* Cambridge: Cambridge University Press.

Yirmiya, N., Sigman, M., Kasari, C., & Mundy, P. (1992). Empathy and cognition in high-functioning children with autism. *Child Development, 63*, 150-160.

Zahn-Waxler, C., Radke-Yarrow, M. , Wagner, E. & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, 28, 126-136.

Zelazo, P., Helwig, C., Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development*, 67, 2478-2492.