Beyond Purity: Moral Disgust toward Bad Character

Roger Giner-Sorolla[1]

Hanah A. Chapman[2]

1. University of Kent

2. Brooklyn College, City University of New York (CUNY)

Running Head: BEYOND PURITY

Corresponding author: Roger Giner-Sorolla, School of Psychology, University of Kent, Canterbury, Kent,. CT 2 7NP, United Kingdom, e-mail rsg@kent.ac.uk

Abstract

Previous studies support a link between moral disgust and impurity, while anger is linked to harm. We challenge this strict correspondence, and show that disgust is sensitive to information about moral character, even for harm violations. By contrast, anger is sensitive to information about actions, including their moral wrongness and consequences. Study 1 examined disgust and anger toward an action that indicates especially bad moral character (animal cruelty) versus an action that is more wrong (domestic abuse). Animal cruelty was associated with more disgust, whereas domestic abuse was associated with more anger. Studies 2 and 3 manipulated character by varying the agent's desire to cause harm, and also varied the action's harmful consequences. Desire to harm predicted only disgust (controlling for anger), while consequences were more closely related to anger (controlling for disgust). Taken together, these results indicate disgust responds to evidence of bad moral character, not just to impurity.

Beyond Purity: Moral Disgust toward Bad Character

Hostile emotions such as disgust and anger play a controversial part in psychological and philosophical questions of moral judgment. Emotionalist accounts (e. g., Haidt, 2001; Kass, 2005) argue that hostile feelings <u>do</u> and, sometimes, <u>should</u> inform moral decisions. But other prescriptive views draw distinctions between emotions; disgust, in particular, is said to have an undesirable effect on judgment due to its irrational nature (e.g., Nussbaum, 2006). Which hostile emotions are felt toward moral violations, and why, have thus become questions of interest.

Research so far has mostly distinguished disgust from anger on the basis of the moral rule violated. In the Community/Contempt—Autonomy/Anger—Divinity/Disgust (CAD) hypothesis (Rozin, Lowery, Imada & Haidt, 1999), anger responds to rights or "autonomy" violations such as harm or unfairness, while disgust responds to "divinity" or "purity" violations that contaminate the body or soul, such as taboo sexual acts (Horberg, Oveis, Keltner & Cohen, 2009; Russell & Giner-Sorolla, 2013). If people say they are "disgusted" at harm or unfairness, it has been argued, this is only semantic confusion with anger (e.g., Nabi, 2002). The idea that different moral rules are associated with different emotions is significant for broader theories of moral judgment, since it supports the notion that there are different kinds or domains of morality (e.g. Haidt, 2001), in contrast to domain-general accounts in which the same cognitive processes support all moral judgments (e.g. Cameron, Lindquist & Gray, 2015).

However, some findings of moral disgust at non-purity offenses are not easily explained away. Lab studies have shown physiological and verbal disgust responses to unfairness, independent from anger (Chapman, Kim, Susskind & Anderson, 2009; Cannon, Schnall, & White, 2011; Moretti & di Pellegrino, 2010). Also, the relationship between endorsement of faces and words relevant to disgust was shown to be significant in scenarios

involving rights violations, even when anger endorsement was statistically controlled (Giner-Sorolla, Gutierrez & Vasiljevic, 2012). "Disgust" toward rights violations, then, is not entirely explicable as a verbal synonym for "anger." In addition to its importance for theories of morality, disgust toward non-purity offenses would also represent a striking example of exaptation, an evolutionary process whereby a preexisting structure assumes a new functional role (Rozin, Haidt & McCauley, 2008). Disgust is believed to have originated as a defense against poison and disease (Rozin et al., 2008), so disgust toward non-purity offenses would constitute a dramatic change in its function.

If moral disgust is not restricted to purity violations, nor just a synonym for anger, what then is its role in morality? Some have speculated that disgust may arise when a person is seen as having a fundamentally bad character (Haidt, 2003; Hutcherson & Gross, 2011; Miller, 1997; Rozin et al., 2008), regardless of what moral norm they have broken. Although this idea has not been tested directly, individuals do predominantly feel shame when others express disgust toward them (Giner-Sorolla & Espinosa, 2011), and appraisals of one's own bad character are linked to shame (Lewis, 2010). Functionally, the avoidance action tendencies and indelible "taint" of disgust (Rozin et al., 2008)   fit well with the enduring and permanent stigma of bad character.

In contrast to disgust, anger may be more sensitive to factors that influence whether a particular act is viewed as right or wrong, such as its consequences. For example, manipulations of whether or not an action caused harm influence anger more than disgust (Gutierrez & Giner-Sorolla, 2007; Russell & Giner-Sorolla, 2011a; Giner-Sorolla, Caswell, Bosson & Hettinger, 2012). As well, individuals primarily feel guilt when others express anger toward them (Giner-Sorolla & Espinosa, 2011), and appraisals of one's own bad actions (rather than character) are linked to guilt (Lewis, 2010). Functionally, because of its approach action tendencies (Carver & Harmon-Jones, 2009), anger is more suited to

correcting a mistaken or negligent harm committed by someone of basically good character (Fischer & Roseman, 2007). For these reasons, the moral wrongness of an act and its harmful consequences may influence anger to a greater degree than disgust.

To test these predictions, we conducted three studies that independently manipulated act- and character-related variables in descriptions of harmful moral violations. In Study 1, we built on previous findings that some actions that are, themselves, relatively less immoral yield stronger inferences of bad character than others that are more immoral (Tannenbaum, Uhlmann, & Diermeier, 2011). In Studies 2 and 3, we manipulated the relative salience of moral character and act judgments by varying the desire to harm others (Aristotle, trans. 1890; Reeder, Kumar, Hesson-MacInnis, & Trafimow, 2002), as well as whether or not the action had harmful consequences. We predicted that information leading to inferences of bad moral character would primarily affect disgust, with a lesser influence on anger. By contrast, information about the moral badness of the act, namely its wrongness and consequences, should primarily affect anger, with a weaker influence on disgust.

Study 1

One situation in which act and character evaluations can be dissociated is when acts that are relatively less immoral nonetheless provide strong signals of poor character (Tannenbaum et al., 2011; Uhlmann, Zhu, & Diermeier, 2014). For example, Tannenbaum and colleagues (2011) measured act and character judgments toward a scenario in which an individual reacts to his girlfriend's infidelity by beating her, and another scenario in which an individual reacts to the same situation by beating his girlfriend's cat. Participants judged that the cat-beater had worse moral character than the woman-beater, even though the act of beating a woman was judged to be more immoral. These findings demonstrate that act and character evaluations can vary independently, but without examining the emotions associated with them. In Study 1 we therefore carried out a pre-registered replication and extension of

Tannenbaum et al. (2011, Study 1a); registration at osf.io/ynvhz/[1]). We presented the original scenarios in which a cat or a woman was beaten, measuring character judgments, wrongness judgments (as an index of act evaluations), anger and disgust. If character judgments are primarily related to disgust, then the cat-beater (who has worse character) should evoke more disgust than the woman-beater. Conversely, if act judgments are primarily related to anger, then the woman-beater (whose actions are more immoral) should evoke more anger than the cat-beater. As well, character judgments should be more strongly correlated with disgust than with anger, whereas act judgments should be more strongly correlated with anger than disgust.

Method

In this article we report all measures in each of our studies, all manipulations, any data exclusions, and the sample size determination rule. In no study was additional data collected after looking at the results.

Participants

One hundred US residents were recruited via Amazon Mechanical Turk, of whom three did not complete the study. This sampling was determined a priori using power analysis, allowing extra recruitment with possible exclusions in mind; 90 cases achieve 80% power to detect a small effect size (d = 0.3) for the critical one-sample t-test comparing responses to the cat- and woman- beater. Of the participants who completed the study, we excluded seven who indicated that the study seemed familiar and three who completed the

---

[1] The registered analysis plan contains one analysis that is not reported here on technical grounds. Our plan for analysis 2c involved an unusual use of residualization, which we believed might allow us to better separate disgust from anger. However, we later realized that this analysis is not statistically meaningful (by definition, residuals average to zero, so it is not informative to compare the average of two residuals).

study in less than half of the median completion time. The final sample size was 87 (34 female, 52 male, 1 other gender identity, mean age = 33.2, range 21-62).

<u>Procedure</u>

Participants were asked to read and evaluate two scenarios (Tannenbaum et al., 2011). The *woman-beater* scenario read as follows: "John learns that his girlfriend of 8 years has been sleeping around with another man. Upon hearing this, John becomes overwhelmed with rage and beats up his girlfriend." The *cat-beater* scenario replaced "beats up his girlfriend" with "beats up his girlfriend's cat".

After being introduced to the scenarios, participants were asked to evaluate John and Robert's actions and their moral character, as well as to report on their emotional reactions to the scenarios both verbally and by endorsing emotional expressions. Act, character, verbal emotion ratings and expression endorsement ratings were presented in separate blocks, in random order. To assess act evaluations, participants indicated whose actions were more immoral, which action should be punished more severely, and which action deserves more blame, on a 1-7 scale (1 = John beating up his girlfriend, 7 = Robert beating his girlfriend's cat). To assess character evaluations, participants used the same scale to indicate who is more sick and twisted, who is more likely to have normal human feelings, who is more likely to be sadistic, who is more screwed up, who is more likely to enjoy other peoples' suffering, who is more likely to feel sorry for a homeless person, who is more likely to help the homeless, and who is more likely to feel empathy for a stranger who is suffering (the last three reversed; Tannenbaum et al., 2011).

Participants also rated their relative feelings of disgust and anger toward the scenarios, both verbally and by endorsing photographs of emotional facial expressions (Gutierrez, Giner-Sorolla, & Vasiljevic, 2011). Verbal and expression endorsement measures

were presented in separate blocks. In the expression endorsement measure, participants saw two sets of photographs taken from the Montreal Set of Facial Displays of Emotion (MSFDE; (Beaupré, Cheung, & Hess, 2000). Each set showed three female posers expressing anger or disgust. Participants were asked to rate who they felt more of the depicted emotion towards, using a 1-7 scale (1 = Definitely feel more of this emotion toward John, 7 = Definitely feel more of this emotion toward Robert). The verbal measures asked, in separate items, which target participants felt more anger, infuriated, outrage, disgust, repulsed, and sickened toward (1-7 scale, 1 = Definitely more toward John, 7 = Definitely more toward Robert; cf. Gutierrez & Giner-Sorolla, 2007).

Finally, participants indicated their age, education, political orientation ("when it comes to politics, I am generally…", 7 point scale, very liberal—very conservative), gender, race and ethnicity.

Results

Manipulation checks: act and character ratings. The eight character ratings formed a reliable scale (α = 0.91) and were averaged together, as were the three act ratings (α = 0.93). Replicating Tannenbaum and colleagues' results, a paired-samples t-test indicated that character ratings (M = 4.31, SD = 1.24) were significantly different from act ratings (M = 3.43, SD =1.89; t[86] = -5.37, p < 0.001, d = 0.57). Follow-up one-sample t-tests using the midpoint of the scale as a test value (since participants compared John to Robert) indicated that the cat-beater's actions were judged to be less wrong than the woman-beater's actions (t[86] = -2.82, p = 0.006, d = 0.32). By contrast, the cat-beater was judged to have worse character than the woman-beater (t[86] = 2.36, p = 0.021, d = 0.25).

Emotion ratings. The facial expression and verbal ratings for disgust formed a reliable scale (α = 0.93), as did the expression and verbal ratings for anger (α = 0.93).

However, these four-item scales correlated with each other at a very high level (r = .80), driven by a very high correlation between the verbal disgust and anger ratings (r = 0.95). By contrast, the facial endorsement items proved more able to distinguish between anger and disgust (r = 0.36). Therefore, we used the original single-item face endorsement measures in these analyses.

As predicted, a paired-samples t-test indicated that relative facial disgust ratings (M = 4.36, SE = 0.21) were significantly different from relative facial anger ratings (M = 3.63, SE = 0.20; t[86] = -3.12, p = 0.002, d = 0.42), indicating that the cat-beater and woman-beater differentially evoked disgust and anger. Follow-up one-sample t-tests against the midpoint of the scale showed trends in the predicted directions, with higher disgust for the cat-beater compared to the woman-beater (t[86] = 1.7, p = 0.088, d = 0.21), and higher anger for the woman-beater compared to the cat beater (t[86] = -1.82, p = 0.072, d = 0.21). See Figure 1.
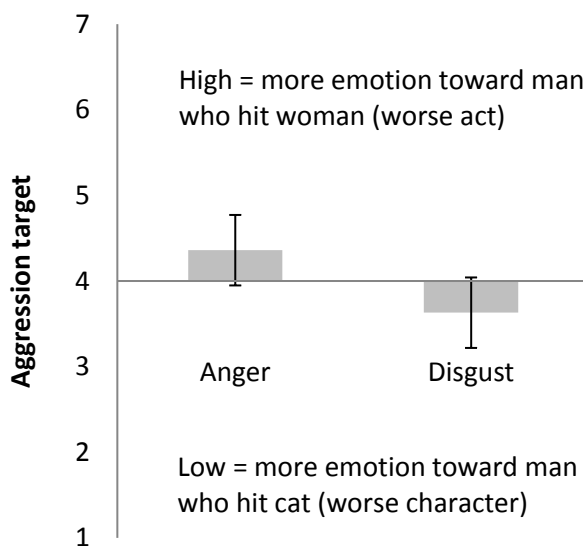

Fig. 1: Means for relative anger and disgust towards the two men in the vignette, Study 1. 95% confidence intervals are shown.

Regression analyses. The relationship between emotion and character ratings was analyzed using a multiple regression with disgust and anger ratings as IVs and character

ratings as the DV. Consistent with our hypotheses, higher disgust was associated with more negative character evaluations ($\beta = 0.38$, $p < 0.001$), while anger was not ($\beta = 0.088$, $p = 0.135$). A t-test for dependent betas indicated that the difference in slopes was significant, $t[84] = 3.07$, $p = 0.0029$.

To examine the relationship between emotion and act ratings, this analysis was repeated with act ratings as the DV. Although we predicted that anger ratings should be associated with more negative act evaluations more so than disgust ratings, we found that both anger ($\beta = 0.46$, $p < 0.001$) and disgust ($\beta = 0.45$, $p < 0.001$) were significantly related to act evaluations. A t-test of the difference between dependent betas indicated that the difference in slopes was not significant ($t[84] = 0.040$, $p = 0.97$).

Study 2

Overall, the results of Study 1 support the hypothesis that disgust is related to judgments of bad moral character, and more equivocally support the idea that anger is related to judgments about the act itself. At the same time, Study 1 did not provide a clean separation of act and character: both the cat-beater and the woman-beater are bad people, and both of their actions are bad. Our Study 2 used a method that more cleanly separates act and character: a paradigm used to study the contribution of desire, consequences, and belief to moral judgments in situations of potential harm (Cushman, 2008). Information about the agent's moral character was varied by describing that he or she either had a *desire* to harm, or no desire to harm, the other person. Desire to harm should primarily influence judgments of moral character and wrongness (Cushman, 2008; Pizarro & Tannenbaum, 2011); we additionally predict it should affect moral disgust more so than anger. Information about the act was varied by manipulating its *consequences,* describing either that the other person was

harmed by the action, or was not. This manipulation, in our predictions, should influence anger more so than disgust. Additionally, to follow Cushman's procedure, we varied whether or not the agent *believed* (or foresaw) that the act would cause harm. This manipulation is not clearly related to moral character (e.g., a good person may not foresee that a particular action will be harmful), nor to the consequences of the act. There is therefore no clear grounds for a prediction about how belief would be related to disgust and anger.

Method

Participants

One hundred and thirty-nine undergraduate students (108 female, 31 male) from a university in southeast England, with a mean age of 20.01 years (ages ranged from 17-38), took part in the study. All participants who completed the online questionnaire during an arbitrary time period were included. This number yielded high power (99.8%) to detect differences between 2 repeated measures means, assuming a correlation of .3 and medium effect size of f = .25; for comparison, the effects of these manipulations on judgments in Cushman (2008) had a conventionally large size, with partial η^2 at .5 and above.

Materials and Procedure

Participants were presented with eight of 64 possible moral scenarios taken verbatim from Cushman's (2008) Study 1, manipulating the presence or absence of harmful belief, desire and consequence in a 2x2x2 design.

These eight factor combinations were counterbalanced with eight scenario settings (e.g., construction site, railway carriage). Participants were randomly assigned to one of eight questionnaires, each with eight scenarios from the set of 64. These were balanced such that each participant saw all eight unique combinations of factors and all eight scenario settings. However, no combination of factors and setting was presented more than once across all questionnaires. The order of scenario presentation was also randomized.

After each scenario, participants were asked to choose one of two picture sets expressing their feelings toward the agent, one showing disgust and one showing anger expressions; each illustration was of three full-intensity black-and-white facial expressions from the MSFDE. They were then asked "How angry are you at [agent]?" and "How disgusted are you at [agent]?" followed by the questions "How wrong was [agent's] behavior?", "How much blame does [agent] deserve?" and "How much should [agent] be punished?" all on seven-point scales. Finally, two questions about moral character were asked, "Do you think that [agent] is mainly a good person or a bad person?" and "Do you think [agent] has good moral standards?" using nine-point scales.

Results

We entered data for each scenario as a separate case, and used the MIXED procedure in IBM SPSS Statistics (v23) to carry out a multilevel analysis predicting each outcome variable from coded independent variables, each representing one of the manipulated traits of the scenarios: that is, desire, belief and consequence (-1 = not present, 1 = present) and their two-way and three-way interactions. We treated the identity of the participant as a random variable, effectively creating a within-participants random-intercept analysis; results of all analyses were substantively similar in significance level and magnitude when full random-slopes models were used.

Our analyses focus on the effects of the manipulations on anger, disgust, and moral character judgments; the Supplementary Online Materials include analyses which generally replicated the effects of our manipulations on the dependent variables of wrongness, blame and punishment from Cushman (2008), and means for the character variable.

Manipulation check: Judgments of moral character. The items "good or bad person" and "good moral standards" were correlated at .91, so they were averaged into one index of moral character, then reversed so that high numbers indicated bad character. As

expected, desire to harm increased judgments of bad character most strongly, $F(1, 966) = 495.85$, $p < .001$, γ coefficient $= 3.13$, with smaller-sized main effects of belief, $F(1, 966) = 95.22$, $p < .001$, $\gamma = 1.55$, and consequence, $F(1, 975.16) = 22.67$, $p < .001$, $\gamma = 0.88$. Desire's effect was moderated further by a significant interaction with belief, $F(1, 966) = 7.48$, $p = .008$, $\gamma = -1.13$, and by the three-way interaction, $F(1, 966) = 6.10$, $p = .014$, $\gamma = 1.08$.

Effects on forced choice of anger vs. disgust expressions. The forced choice item was coded anger $= 0$ and disgust $= 1$. Figure 2 gives the pattern of choices according to desire and consequences. The additional effect of belief, although not theoretically focal, is visualized in Figure S1 in the Supplementary Materials.

Whether the agent desired harm was the strongest predictor of emotion choice, $F (1, 966) = 65.79$, $\gamma = 0.31$, $p < .001$; harmful desire led to more choice of "disgust" over "anger." Belief also predicted choice of disgust over anger, though less strongly, $F(1, 966) = 12.84$, $\gamma = 0.12$, $p = .001$. Harmful consequences, however, did not predict emotion choice overall, $F (1, 966) = 1.74$, $\gamma = 0.04$, $p = .19$. There was also a Desire x Consequences interaction, $(1, 966) = 4.31$, $\gamma = -0.16$, $p = .038$. When there was desire to harm, disgust was chosen more often over anger when there were no consequences than when there were, simple $F (1, 966) = 5.71$, $p = .02$. When there was no desire to harm, consequences had no effect on choice, simple $F (1, 966) = 0.28$, $p = .60$.
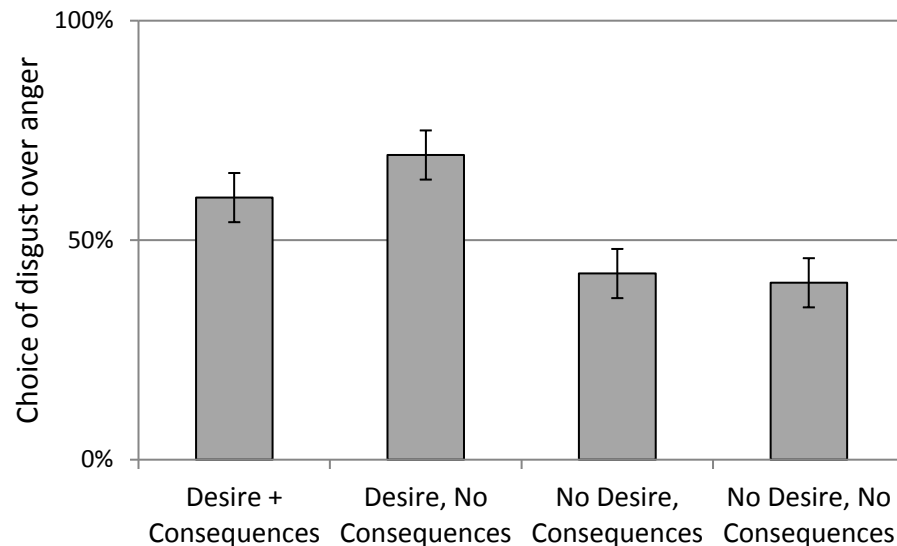
Fig. 2. Percent of participants choosing disgust over anger in Study 2, with 95% confidence intervals.

Anger, controlling for disgust. Since there was considerable shared variance between ratings of anger and disgust ($r = 0.74$), we analyzed the effects of desire, belief and consequences on anger with disgust as a covariate. Results for desire and consequences are shown in Figure 3; see Figure S2 for a visualization of the effect of belief. Unadjusted results are shown in Figure S3. This analysis found significant main effects on anger from belief, $F(1, 1008.92) = 27.21$, $\gamma = 0.58$, $p < .001$, and consequences, $F(1, 972.82) = 57.89$, $\gamma = 0.52$, $p < .001$, but not desire, $F(1, 1045.55) = 1.59$, $\gamma = 0.36$, $p = .207$[2]. Also, the Belief x Desire interaction was significant, $F(1, 961.38) = 12.07$, $\gamma = -0.31$, $p < .001$; desire to harm did not increase anger when the agent believed the act to be harmful (no desire adjusted $M = 4.52$, desire adjusted $M = 4.37$) but did increase anger when the agent did not believe the act to be harmful (no desire adjusted $M = 3.85$, desire adjusted $M = 4.21$). No other interactions were significant.

---

[2] Non-integer degrees of freedom are a consequence of the SPSS MIXED procedure using Satterthwaite approximations to the F distribution when a covariate is included. (SPSS Inc. 2005)

Disgust, controlling for anger. Reversing the role of the two emotions from the anger analysis, so that disgust was the DV and anger the covariate, a different pattern of effects emerged (Figure 3; see Figure S2 for effects of belief and Figure S3 for unadjusted results). Desire was the strongest predictor of disgust, $F(1, 1025.79) = 209.29$, $\gamma = 1.51$, $p < .001$, and belief was a significant but weaker predictor, $F(1, 1019.78) = 52.21$, $\gamma = 0.70$, $p < .001$. Consequences did not predict disgust, $F(1, 1001.30) = 0.71$, $\gamma = 0.28$, $p = .401$. The Desire x Consequence interaction was weak but significant, $F(963.30) = 4.39$, $\gamma = -0.50$, $p = .036$. When there were no harmful consequences, the effect of desire in increasing disgust was slightly stronger (adjusted Ms = 5.06 vs. 3.68) than when there were harmful consequences (adjusted Ms = 4.96 vs. 3.91).

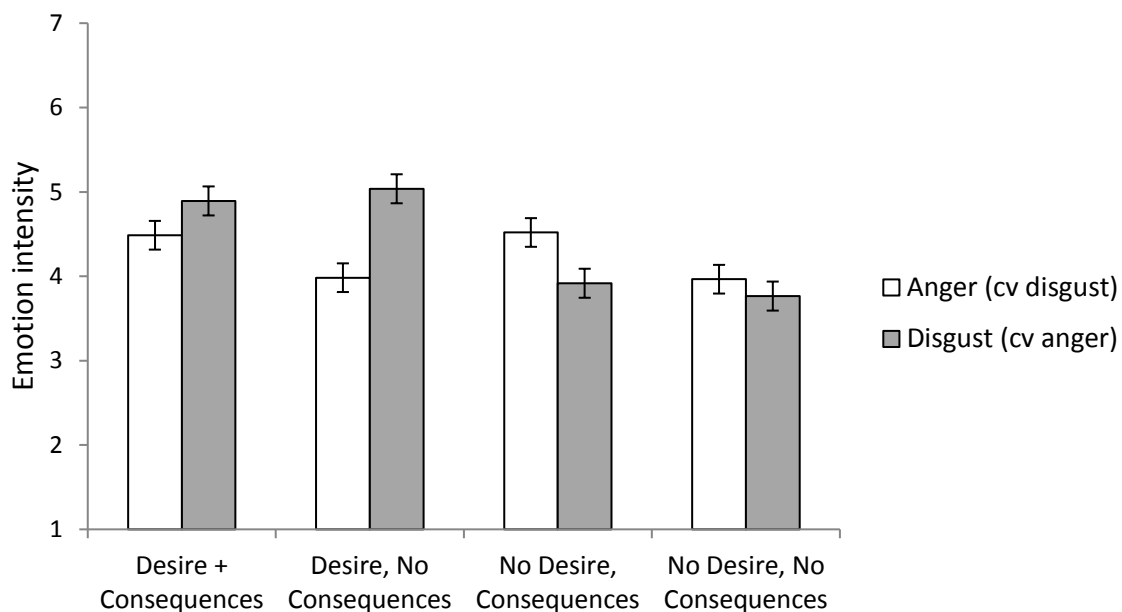Fig. 3. Adjusted means for verbal ratings of anger and disgust by Desire and Consequences conditions, Study 2, covarying the other emotion, with 95% confidence intervals.

Overall, the adjusted means of anger and disgust reflected these main effects of desire increasing disgust (more than anger) and of consequences increasing anger (more than disgust). Disgust predominated over anger when there was desire to harm, and most strongly

so when there were no consequences; anger predominated over disgust when there were consequences, but no desire to harm.

Mediation of desire-disgust link by character judgment. We tested the ability of character judgment to account for the effect of our manipulation of harmful desires on disgust, treating moral character (not reversed) as a mediator, the desire manipulation as the predictor and verbal disgust ratings as the outcome, with verbal anger ratings as a covariate. We performed multilevel mediation analyses (see Preacher, Zyphur, & Zhang, 2010), estimating the indirect path through a bootstrap procedure with 5000 resamples. This showed moral character to be a significant partial mediator between desire and disgust; when accounted for, the direct path between desire and disgust remained significant (Figure 4).

A "reversed" mediation with character as outcome and disgust as mediator gave an indirect path effect of 0.36 (95% CI: 0.24 to 0.47), which, however, did not completely account for the direct effect of desire manipulation on character, direct effect coefficient = 1.54 (95% CI: 1.30 to 1.77), total effect = 1.89. The relative strength of these paths, however, should be interpreted cautiously because factors such as relative reliability can interfere with their comparison; in general, a mediation model can only be suggestive about the causal ordering of variables which could influence each other and were measured at approximately the same time (Thoemmes, 2015; Lemmer & Gollwitzer, 2016). The safest interpretation is that there is a mutual influence of disgust and character which helps to explain the effect of the manipulation of desire on both variables.
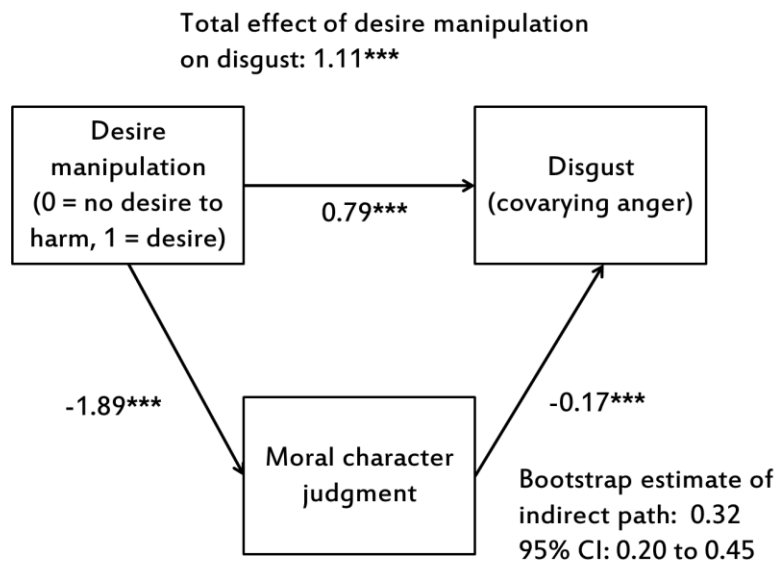
Total effect of desire manipulation
on disgust: 1.11***

```
┌─────────────────┐                          ┌─────────────────┐
│     Desire      │                          │                 │
│  manipulation   │        0.79***           │    Disgust      │
│ (0 = no desire to│ ──────────────────────> │ (covarying anger)│
│ harm, 1 = desire)│                          │                 │
└─────────────────┘                          └─────────────────┘
     -1.89***                                      -0.17***
        ┌─────────────────┐
        │ Moral character │     Bootstrap estimate of
        │    judgment     │     indirect path:  0.32
        └─────────────────┘     95% CI: 0.20 to 0.45
```

Fig. 4. Mediation analysis: Effect of desire manipulation on disgust via character, Study 2.
*** = p > .001.

Study 3

A limitation of the first two studies was a within-participants design, which may have allowed participants to explicitly compare scenarios. Study 3 therefore used a between-participants design. The results of Study 2 also suggested a number of simplifications.

We wanted to focus this design on the most distinctive factors separating anger and disgust. First, we removed the manipulation of belief, giving information only about harmful desire and harmful consequences. Belief seemed to relate to both anger and disgust in about equal measure, unlike harmful consequences which exclusively predicted anger and desire which exclusively predicted disgust. Also, we further simplified the design into only desire / no consequences and no desire / consequences conditions, the two conditions that produced the clearest differences between anger and disgust ratings in Study 2.

With these reductions, we could accommodate a new manipulation based on an observation about our measures in Studies 1 and 2: the emotion items had been targeted on

the person committing the act. To see whether this context of person-focus was necessary to strengthen the effects of desire and character on differentiating disgust from anger, we asked emotion questions in Study 3 either about the person, or about the act. Because desire and disgust showed a connection to person-based judgments of moral character in Studies 1 and 2, we thought that the evaluation of the person would show a stronger influence of disgust, while the evaluation of the act was conceptually related more to its consequences, and so would show a stronger influence of anger.

Method

Participants

One hundred and sixty US residents recruited via Amazon Mechanical Turk took part in the study. This number was determined a priori and gives 88% power to detect a medium effect size (f = .25) between two groups, taking into account the medium to large key effects in Study 2; participants who did not begin the questionnaire after signing up were excluded and new ones recruited until the target number (160) was reached. Two participants were further excluded for failing to answer the check questions (see below) correctly, leaving a total of 158 (110 male, 48 female), with a mean age of 31.02 years (range 19-72).

Materials and Procedure

The scenarios used were identical to Study 2, except that the information about the agent's beliefs was removed, so that only desire and consequences were manipulated; and only the "Consequence, No Desire" (henceforth "Consequence") and "Desire, No Consequence" (henceforth "Desire") versions of the stories were used. The design was between participants, 2 (Consequence vs. Desire condition) x 2 (Emotion measure, Person vs. Act focused).

Each participant was randomly assigned to read one of the scenarios in either the Consequences or the Desire version. After reading the scenario, they completed emotion

measures, aimed at the actions of the target in the act conditions (e.g., "How ANGRY are you at the actions of Amy?"), or anger or disgust at the target in the person conditions (e.g., "How ANGRY are you at Amy?"). As in Study 2, participants made a dichotomous choice between disgust and anger expressions, and also provided an expression-based scale rating in which they were asked how much they felt like the disgust and anger faces (1-7 scale, "not at all" to "very"). Participants also completed additional verbal measures of emotions as in Study 1 (cf. Gutierrez & Giner-Sorolla, 2007); the ones relevant to anger were "outraged," "furious" and "angry" while the ones relevant to disgust were "disgusted", "revolted" and "sickened" (1-7 scale, "not at all" to "very"). Three additional words measuring compassion, kindness and hatred were also rated but are not relevant to the present hypotheses.

After the emotion questions, further judgments were assessed in a randomly presented block, without varying the wording according to the act/person conditions, which were intended to apply only to the emotion measures. Participants completed the same character-related judgments as in Study 2 ("good /bad person" and "moral standards", on nine-point scales which were transformed before analysis) as well as two more character-related items ("How moral/how trustworthy is [agent]"?, on seven-point scales). Participants also completed a number of items regarding the act's wrongness, the responsibility of the actor, whether they should be punished, and the appropriateness of various legal actions toward him/her. These items are not relevant to the current hypotheses; the full item descriptions and results are given in the Supplemental Online Materials. Finally, two checks on attention were included: one asking whether a character named Tim appeared in the scenario (he never did) and another asking whether anyone was harmed in the vignette (correct answer varying according to condition).

Results

The analysis of emotion choices used a 2 x 2 ANOVA between-subjects design (Consequence/ Desire x Act/Person emotion focus), emotion ratings used a 2 x 2 x 2 ANOVA mixed design (Consequence/Desire x Act/Person emotion focus x Anger/Disgust, within) and all other scores were analyzed in a one factor (Consequence/Desire) between-subjects design.

Manipulation check: Judgments of moral character. The two nine-point items measuring moral character were rescaled, multiplying each by 7/9, to correspond to the seven-point items. Low scores indicated worse moral character. The four items formed a reliable scale at $\alpha = .95$. Participants in the Desire condition attributed worse moral character to the agent (M = 1.91, SD = 1.00) than those in the Consequence condition (M = 4.22, SD = 1.40); $F(1, 156) = 140.44$, $p < .001$, $\eta^2_p = .47$.

Effects on forced choice of anger vs. disgust expressions. The dichotomous choice was coded as 0 = anger, 1 = disgust, so that the means represent the proportion choosing disgust (Figure 5). A significant effect of Consequence/Desire showed more choice of disgust in the Desire condition (.71, SD = .46) than in the Consequence condition (.45, SD = .50), $F(1, 154) = 11.02$, $p = .001$, $\eta^2_p = .07$. Act/Person had no significant main effect, $F < 1$, or interaction with Consequence/Desire, $F(1, 154 = 2.69, p = .104, \eta^2_p = .02)$. However, the simple effect of Consequence/Desire on anger/disgust choice was significant when the emotions focused on the person (Desire, M = .74, SD = .44; Consequence M = .37, SD = .49, simple effect significant, p = .001) but not when they focused on the act (Desire M = .67, SD = .48; Consequence M = .54, SD = .51, simple effect ns, p = .239).
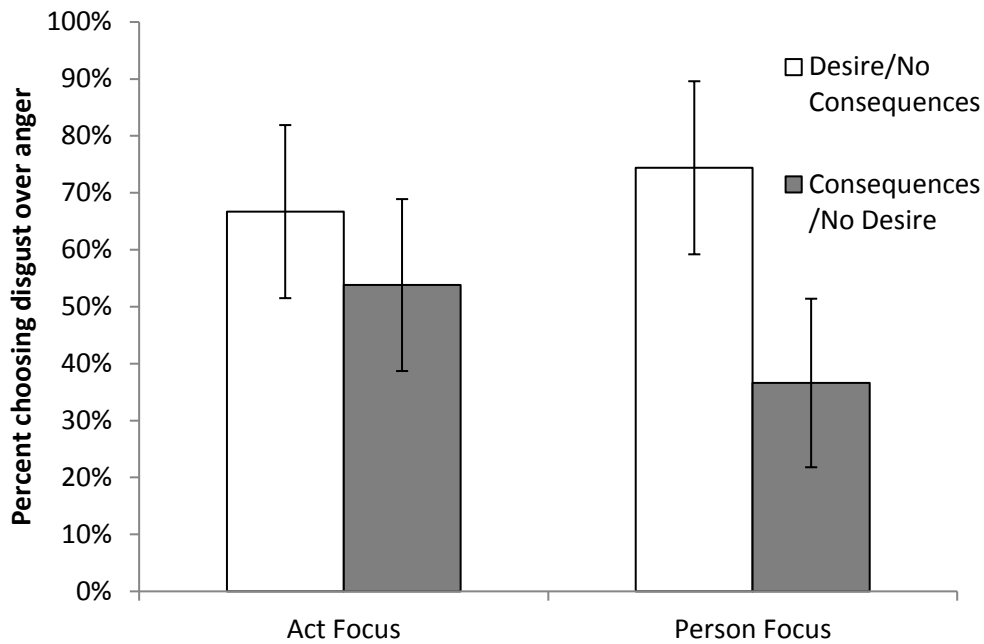
Fig. 5. Percent of participants choosing disgust over anger, Study 3, with 95% confidence intervals.

Scaled ratings of emotion. The scaled facial expression anger item and the three

additional verbal anger items formed a scale with reliability $\alpha = .95$, and likewise the four

disgust-related items formed a scale with reliability $\alpha = .95$. However, these four-item scales

correlated with each other at a very high level ($r = .80$), likely driven by the high correlation

between the verbal disgust and anger ratings ($r = .81$), whereas the scaled facial endorsement

items proved more able to distinguish between anger and disgust, although there was still a

good deal of overlap ($r = .66$). Therefore, we used the two single-item scaled face

endorsement measures in these analyses, reporting the three-item verbal scale analyses

separately only when they differed in level of significance from the single-item. Figure S5

gives unadjusted means for face endorsement by condition.

The predicted interaction between emotion measure (Anger/Disgust) and

Consequence/Desire was significant, $F(1,154) = 14.14$, $p < .001$, $\eta^2_p = .08$, and was in turn

qualified by a marginal three-way interaction also involving Act/Person, $F(1,154) = 3.62$, $p =$

.06, $\eta^2_p = .02$ (on the verbal measures, the three-way interaction was significant, $F(1, 154 = 5.48, p = .02, \eta^2_p = .03$). To further interpret the pattern of these interactions, we used covariance analyses as in Study 2, applying the Consequence/Desire x Act/Person model to disgust with anger as a covariate, and to anger with disgust as a covariate (Figure 6).
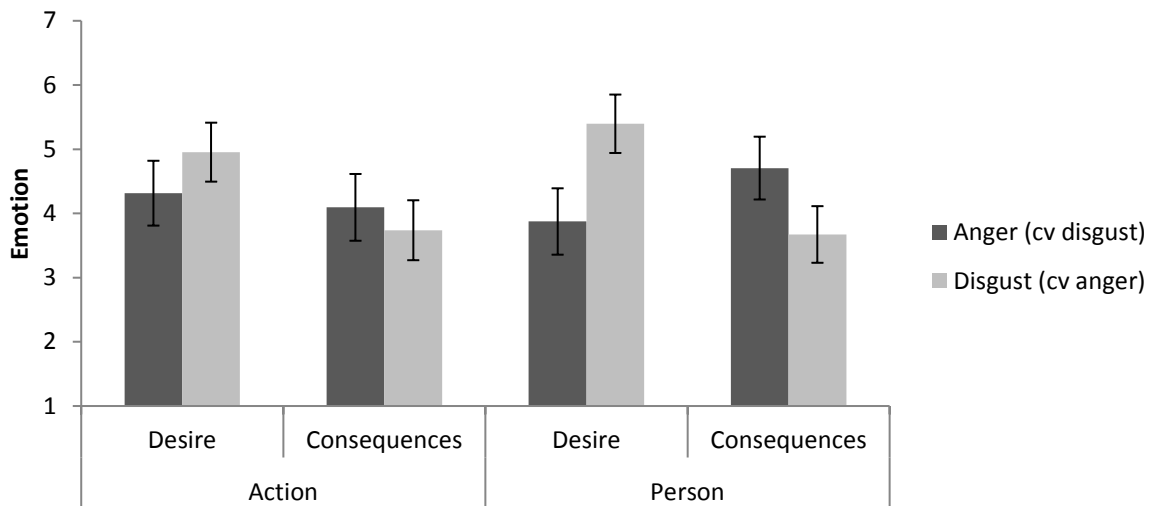

Fig. 6. Means of endorsement of facial anger and disgust, covarying the other emotion, by condition, Study 3, with 95% confidence intervals.

For facial disgust as a DV, the Consequence/Desire main effect was the only significant factor, $F (1, 153) = 38.48, p<.001, \eta^2_p = .20$, with no interaction involving Act/Person, $F (1, 153) = 1.22, p= .271, \eta^2_p =.008$. Simple effects showed that disgust (controlling for anger) was higher in the Desire condition compared to the Consequence condition, within both the Act focus condition, $F(1,153) = 12.88, p < .001, \eta^2_p = .08$, and the Person focus condition, $F(1, 153) = 28.82, p< .001, \eta^2_p = .16$. Using verbal disgust instead, the interaction with Act/Person became significant, $F (1,153) = 5.66, p = .019, \eta^2_p = .04$. The disgust gap between Desire and Consequence conditions was larger in the Person condition, $F(1,153) = 49.39, p < .001, \eta^2_p = .23$, than in the Act condition, $F(1,153) = 11.06, p = .001, \eta^2_p = .07$. Overall, disgust was higher when the agent desired negative consequences but none occurred, compared to when negative consequences occurred but the agent did not desire them. For the

verbal ratings, this effect was stronger in the conditions where emotions were focused on the persons rather than the act, supporting the interpretation that disgust at mere desire to harm indicates a judgment of personal character.

For anger as a DV, although the main effects of Consequence/Desire and Act/Person were not significant (p > .25), their interaction was, $F (1, 153) = 4.57$, $p = .034$, $\eta^2_p = .03$. The simple effect of Consequence/Desire was significant within the Person condition, $F (1, 153) = 4.97$, $p = .027$, $\eta^2_p = .03$; but not within the Act condition, $F < 1$. For person judgments, anger covarying disgust was higher in the Consequence condition (adjusted M = 4.71, SE = 0.25) than in the Desire condition (adjusted M = 3.87, SE = 0.26). For act judgments, the means went in the other direction (Consequence adjusted M = 4.09, SE = 0.26; Desire adjusted M = 4.32, SE = 0.26). Using verbal ratings instead, the only change in significance patterns was that the main effect of Consequence/Desire was now significant if only by a slim margin, $F (1,153) = 4.12$, $p = .044$, $\eta^2_p = .03$; anger controlling for disgust was higher in the Consequence condition (M = 4.20, SE = .14) than in Desire (M = 3.75, SE = .15). These results indicate that anger was higher when negative consequences occurred even though the agent did not desire them, but contrary to our hypotheses, only when emotions were focused on the person rather than the action.

Mediation of desire-disgust link by character judgment. As in Study 2, a mediation analysis was carried out to establish the link between character and disgust in the context of the manipulation of desire vs. consequences. The analysis collapsed across both person and act conditions, with character as a mediator between the manipulation and the scaled facial measure of disgust, and covarying anger. This showed a significant indirect path, but this time, no direct path, indicating full mediation of effects on disgust by character (Figure 7). For verbal measures of disgust, controlling for verbal measures of anger, mediation was also significant (95% CI = 0.34 to 1.04) but only partial, as a significant direct path from the

manipulation to disgust remained (b =0.61, p =.006). When anger controlling for disgust was the DV, however, there was no significant mediation by character.

A "reversed" mediation as in Study 2 with character as outcome and disgust as mediator gave an indirect path effect of 0.47 (95% CI: 0.22 to 0.81) that, unlike the disgust-outcome mediation, did not completely account for the direct effect of condition on character, direct effect coefficient =1.35 (95% CI: 0.96 to 1.73). Again, bearing in mind that the experimental design did not allow clear distinction of the relative causal role of disgust and character, the best interpretation of these analyses is that character and disgust were involved with each other's effects under the manipulation of consequences and desire.
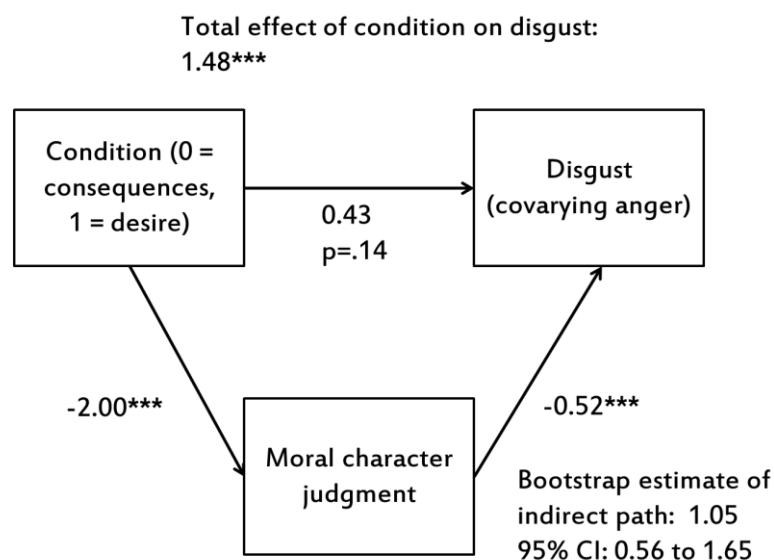
Total effect of condition on disgust:
1.48***

```
┌─────────────────┐                           ┌─────────────────┐
│  Condition (0 =  │                           │    Disgust       │
│  consequences,   │      0.43  ───────────►   │ (covarying anger)│
│   1 = desire)    │      p=.14                │                  │
└─────────────────┘                           └─────────────────┘
         \                                              ↗
    -2.00***                                        -0.52***
           \        ┌─────────────────┐        /
            ──────► │  Moral character │ ──────       Bootstrap estimate of
                    │     judgment     │              indirect path:  1.05
                    └─────────────────┘              95% CI: 0.56 to 1.65
```

Fig. 7. Mediation analysis: Effect of desire manipulation on disgust via character, Study 3.

*** = p > .001.

Discussion

The results of three studies supported our hypotheses that disgust would respond, more so than anger, to information about character. In each study, even immoral acts not related to impurity elicited disgust more than anger if taken as evidence for the actor's evil character.

These results are not just due to semantic confusion of the word "disgust" with feelings of anger, because they were obtained with measures giving both emotions as options, and statistically controlling for the other emotion.

We found more limited support for the proposal that anger more so than disgust would respond to act-related variables, including moral wrongness and harmful consequences. In particular, Study 3 found that anger was more sensitive to consequences vs. character only when emotions were focused on the person rather than the act. For emotions felt directly toward unintended harmful acts, anger and disgust were chosen with approximately equal frequency, and rated at approximately equal intensity. Thus, the distinction between anger and disgust appears clearer when persons are the target of the emotion.

Overall, our results suggest that disgust and anger are sensitive to different moral variables, at least under certain conditions. These findings are only relative: each study showed strong co-activation of disgust and anger. However, co-activation of disgust and anger does not mean that they are the same thing. Further work might explore the different social functions of these two reactions, and in particular the tendency of disgust to encourage withdrawal behavior with a relatively low cost, appropriate to someone who is seen as having irredeemably bad character, compared to the potentially costly approach motivation linked to anger (Carver & Harmon-Jones, 2009). Indeed, allowing participants the opportunity to act on their feelings could amplify differences between disgust and anger, much as the opportunity to act increases anger-related cortical asymmetries (Carver & Harmon-Jones, 2009). For now, our results have shown how it is possible that any kind of immoral behavior – acting unfairly, betraying one's group, and so on – might lead to more disgust than anger, if the act is attributed to a "rotten" moral character.

Author Contributions

R. Giner-Sorolla developed the concept and design for Studies 1-3, together with H. Chapman for Study 1. Data were collected by H. Chapman for Study 1 and under the supervision of R. Giner-Sorolla for Studies 2-3. Analyses and manuscript writing were conducted by R. Giner-Sorolla and H. Chapman. Both authors approved the final version of the manuscript for submission.

References

Aristotle (trans. 1890). *The ethics of Aristotle: With introductory essay by George Henry-Lewes* (D. R. Chase, trans.) London: Walter Scott Publishing Co. doi: 10.1037/12981-000

Beaupré, M. G., Cheung, N., & Hess, U. (2000). The Montreal Set of Facial Displays of Emotion [Slides]. Available from Ursula Hess, Department of Psychology, Humboldt University Berlin.

Cannon, P. R., Schnall, S., & White, M. (2011). Transgressions and expressions: Affective facial muscle activity predicts moral judgments. *Social Psychological and Personality Science*, *2*(3), 325–331. doi:10.1177/1948550610390525

Cameron, C. D., Lindquist, K. A., & Gray, K. (2015). A constructionist review of morality and emotions: No evidence for specific links between moral content and discrete emotions. *Personality and Social Psychology Review*, 19(4), 371–394. doi: 10.1177/1088868314566683

Carver, C. S., & Harmon Jones, E. (2009). Anger is an approach-related affect: evidence and implications. *Psychological Bulletin*, 135(2), 183–204. http://doi.org/10.1037/a0013965

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, *323*(5918), 1222. doi:10.1126/science.1165565

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*(2), 353-380. doi: 10.1016/j.cognition.2008.03.006

Fischer, A. H., & Roseman, I. (2007). Beat them or ban them: The characteristics and social functions of anger and contempt. *Journal of Personality and Social Psychology*, *93*, 103–115. http://dx.doi.org/10.1037/0022-3514.93.1.103

Giner-Sorolla, R., & Espinosa, P. (2011). Social cueing of guilt by anger and shame by

disgust. *Psychological Science, 22,* 49-53. doi: 10.1177/0956797610392925

Giner-Sorolla, R., Caswell, T. A., Bosson, J., & Hettinger, V. (2012). Emotions in sexual

morality: Testing the separate elicitors of anger and disgust. *Cognition and Emotion,

26*, 1208-1222. doi: 10.1080/02699931.2011.645278

Gutierrez, R., Giner-Sorolla, R., & Vasiljevic, M. (2012). Just an anger synonym? Moral

context influences predictors of disgust word use. *Cognition and Emotion, 26*(1), 53-64.

doi: 10.1080/02699931.2011.567773

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to

moral judgment. *Psychological Review*, *108*(4), 814–834. doi: 10.1037//0033-295X

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith

(Eds.), *Handbook of Affective Sciences* (pp. 852–870). New York: Oxford University

Press.

Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of

purity. *Journal of Personality and Social Psychology, 97*(6), 963-976. doi:

10.1037/a0017423

Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social-functionalist account

of anger, disgust, and contempt. *Journal of Personality and Social Psychology, 100*(4),

719–737. doi: 10.1037/a0022408

Kass, L.R. (2005). The wisdom of repugnance: Why we should ban the cloning of humans.

Valparaiso University Law Review, 32, 679-706.

Lemmer, G., & Gollwitzer, M. (in press). The "true" indirect effect won't (always) stand up:

When and why reverse mediation testing fails. *Journal of Experimental Social

Psychology*. Retrieved from http://www.sciencedirect.com

/science/article/pii/S0022103116302876. doi:10.1016/j.jesp.2016.05.002

Lewis, M. (2010). Self-conscious emotions: Embarrassment, pride, shame and guilt. In M. Lewis, J. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of Emotions* (3rd ed., pp. 742–756). Guilford.

Miller, W. I. (1997). *The anatomy of disgust*. Cambridge, MA: Harvard University Press.

Moretti, L., & Di Pellegrino, G. (2010). Disgust selectively modulates reciprocal fairness in economic interactions. *Emotion*, *10*(2), 169–180. doi: 10.1037/a0017826

Nabi, R. L. (2002). The theoretical versus the lay meaning of disgust: Implications for emotion research. *Cognition and Emotion, 16*(5), 695-703. doi: 10.1080/02699930143000437

Nussbaum, M. (2006). *Hiding from humanity: Disgust, shame, and the law.* Princeton, NJ: Princeton University Press.

Ortony, A., Clore, G. L., Collins, A. (1990). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Pizarro, D.A. & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. Shaver, (Eds.), *The social psychology of morality: Exploring the causes of good and evil.* (pp. 91–108). Washington, DC: American Psychological Association.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*, 209-233. doi:10.1037/a0020141

Reeder, G. D., Kumar, S., Hesson-McInnis, M., & Trafimow, D. (2002). Inferences about the morality of an aggressor: The role of perceived motive. *Journal of Personality and Social Psychology, 83*(4), 789-803. doi: 10.1037/0022-3514.83.4.789

Rozin, P., Haidt, J., & McCauley, C. R. (2008). Disgust. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of Emotions* (3rd ed., pp. 757–776). New York: Guilford.

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. doi: 10.1037/0022-3514.76.4.574

Russell, P. S., & Giner-Sorolla, R. (2011). Moral anger, but not moral disgust, responds to intentionality. *Emotion, 11*(2), 233-240. doi: 10.1037/a0022598

Russell, P. S., & Giner-Sorolla, R. S. (2013). Bodily moral disgust: What it is, how it is different from anger, and why it is an unreasoned emotion. *Psychological Bulletin, 139*(2), 328. doi: 10.1037/a0029319

SPSS Inc. (2005) . Linear mixed-effects modelling in SPSS: An introduction to the MIXED procedure. Retrieved from http://www.spss.ch/upload/1126184451_Linear%20Mixed%20Effects%20Modeling%20in%20SPSS.pdf.

Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology, 47*(6), 1249–1254. doi: 10.1016/j.jesp.2011.05.010

Thoemmes, F. (2015). Reversing arrows in mediation models does not distinguish plausible models. *Basic and Applied Social Psychology*, *37*(4), 226–234. http://doi.org/10.1080/01973533.2015.1049351

Uhlmann, E. L., Zhu, L., & Diermeier, D. (2014). When actions speak volumes: The role of inferences about moral character in outrage over racial bigotry. *European Journal of Social Psychology, 44(*1), 23–29. doi: 10.1002/ejsp.1987