

## Original Article

### To Defer or To Stand Up? How Offender Formidability Affects Third Party Moral Outrage

Niels Holm Jensen, Department of Psychology, Aarhus University, Aarhus, Denmark. Email: [nielshje@psy.au.dk](mailto:nielshje@psy.au.dk) (Corresponding author).

Michael Bang Petersen, Department of Political Science, Aarhus University, Aarhus, Denmark.

**Abstract:** According to models of animal behavior, the relative formidability of conspecifics determines the utility of deferring versus aggressing in situations of conflict. Here we apply and extend these models by investigating how the formidability of exploiters shapes third party moral outrage in humans. Deciding whether to defer to or stand up against a formidable exploiter is a complicated decision as there is both much to lose (formidable individuals are able and prone to retaliate) and much to gain (formidable individuals pose a great future threat). An optimally designed outrage system should, therefore, be sensitive to these cost-benefit trade-offs. To test this argument, participants read scenarios containing exploitative acts (trivial vs. serious) and were presented with head-shot photos of the apparent exploiters (formidable vs. non-formidable). As predicted, results showed that, compared to the non-formidable exploiter, the formidable exploiter activated significantly more outrage in male participants when the exploitative act was serious. Conversely, when it was trivial, the formidable exploiter activated significantly less outrage in male participants. However, these findings were conditioned by the exploiters' perceived trustworthiness. Among female participants, the results showed that moral outrage was not modulated by exploiter formidability.

**Keywords:** moral outrage, formidability, exploitation, third party, trustworthiness

---

## Introduction

Over the course of human evolutionary history, our ancestors have repeatedly been confronted with exploiters – individuals trying to harvest benefits at the expense of others. As demonstrated by a number of laboratory and field studies, this adaptive problem has selected for cognitive and motivational counter-mechanisms including activation of moral outrage and punitive sentiments in second and third parties in response to acts of exploitation (Fehr and

Gächter, 2002; Greene and Haidt, 2002; Kahneman and Frederick, 2002; Price, Cosmides, and Tooby, 2002; Wilson and O’Gorman, 2003). Importantly, however, all exploitative acts are not sanctioned equally. For example, research shows that kin, physically attractive, socially valuable and trustworthy individuals are treated more lenient when engaging in acts of exploitation (Ahola, Christianson, and Hellström, 2009; Feather and Atchison, 1998; Lieberman and Linke, 2007; Petersen, Sell, Tooby, and Cosmides, 2010; Stewart, 1980). In this paper, we add to this research agenda by investigating how the formidability of exploiters is factored into third party reactions to exploitation.

According to the concept of resource holding power (Parker, 1974) derived from evolutionary biology of animal conflict (Archer, 1988; Hammerstein and Parker, 1982; Maynard Smith and Price, 1973), the relative formidability (the capacity to inflict costs on others) of conspecifics determines the utility of deferring versus aggressing in situations of conflict. If an individual is more formidable than oneself, one is better off deferring and, conversely, if an individual is equal or less formidable than oneself, one is better off aggressing. This model has been validated across a range of non-human animal species (Archer, 1988; Dugatkin, 1997; Huntingford and Turner, 1987), just as recent studies of humans have produced results consistent with the model. First, studies have demonstrated that humans across cultures are very adept at gauging physical strength and aggressiveness by using a number of subtle cues such as facial and vocal features (Carré, McCormick, and Mondloch, 2009; Sell, Cosmides, Tooby, Sznycer, von Rueden, and Gurven, 2009; Sell et al., 2010). Second, formidability has the predicted behavioral effects on the self. Hence, an association between physical size and dominance has been established in children and in adolescent boys (Pellegrini et al., 2007; Tremblay et al., 1998) as well as in adult men (Archer and Thanzami, 2007, 2009; Felson, 1996; Gallup, White, and Gallup Jr, 2007). Furthermore, Sell, Tooby, and Cosmides (2009) found that men’s weight lifting ability predicts feelings of entitlement, proneness to anger, history of fighting, utility of personal aggression, and finally, success in conflict.

Importantly, these previous studies on humans have all focused on how the formidability of the self shapes behavior. Fewer studies have focused on the role of others’ formidability in situations of exploitation. The implication of traditional models of animal behavior is that relative formidability matters and that the effects of others’ formidability simply mirror the effects of own formidability (see Archer and Benson, 2008; Cummins, 1999, 2005). Yet, while this might apply to most non-human species, it is not necessarily so in the human animal (Boehm, 2000). Rather, a dissection of the selection pressures involved in the presence of formidable exploiters suggests that we may – under specific conditions – be psychologically designed to react with *greater* outrage in the face of formidable relative to non-formidable exploiters.

Selection for psychological mechanisms designed to produce moral outrage when exploitive acts have occurred is predominantly governed by the adaptive problem of hindering further exploitation (Petersen et al., 2010). Accordingly, in each instance, these mechanisms should be designed to carefully modulate their output by weighing the potential costs of future exploitation against the potential costs of aggressing against the exploiter here and now. In this case, formidable exploiters provide an interesting mix of cross-cutting pressures. On the one hand, as argued above, aggressing against a formidable individual is a risky strategy. On the other hand, a formidable exploiter poses a far greater future threat than a non-formidable exploiter. As argued above, past research shows that formidability

increases exploitive tendencies, and thus a formidable exploiter could be more likely than a non-formidable to repeat an offense. Furthermore, given the larger physical strength, a formidable exploiter is more capable of acting based on a given exploitive tendency. These observations imply that formidable exploiters, on average, will inflict greater future costs than non-formidable exploiters.

The first selective gradient implies that it is adaptive to defer to formidable exploiters. The second selective gradient implies it is particularly adaptive to punish formidable exploiters – especially in the face of serious exploitive acts. For most non-human animals, the first selective gradient dominates because they cannot buffer against the costs of punishing formidable opponents. As argued by Boehm (1999; 2000), however, this does not necessarily apply to humans, given their extraordinary abilities to build and mobilize coalitions. By forming large coalitions, non-formidable humans are able to buffer against the costs associated with aggressing against formidable exploiters. In line with this, hunter-gatherer-bands across continents view alpha-male-like dominant behavior (e.g., rape, theft, unprovoked aggression, undue competition) as deviant (Boehm, 1993; Cashdan, 1980; Furer-Haimendorf, 1967; Gardner, 1991) and broad coalitions from below have been observed to take pre-emptive measures to maintain social control in the group through sanctions. In particular, people have been observed to be vigilant and attentive towards controlling high-power individuals (Boehm, 1999; Diamond, 1997; see also Overbeck, Tiedens, and Brion, 2006).

In this respect, moral outrage seems to be of fundamental importance. Moral outrage is a particular aversive arousal state experienced when others violate social norms, encompassing specialized cognitive (harsh character attributions), affective (anger, contempt, and disgust), as well as behavioral (support for infliction of punitive damages) elements (Darley and Pittman, 2003; Haidt, 2003; Tetlock, Kristel, Elson, Green, and Lerner, 2000). A number of studies have demonstrated that moral outrage fuels punitive sanctions towards norm violators (Fehr and Gächter, 2002; Greene and Haidt, 2002; Kahneman and Frederick, 2002; Price, Cosmides, and Tooby, 2002; Wilson and O’Gorman, 2003). In addition, the activation of outrage also involves more social functions. As Tooby and Cosmides argue (2010; see also Tooby, Cosmides, and Price, 2006), broadcasting moral outrage coordinates the attention of coalitional members towards a common problem, thereby allowing for collective actions. Similarly, Tetlock et al. (2000) argues that moral outrage facilitates meta-norm enforcement; that is, support for punishment of those who do not punish the initial transgressor. In the face of formidable exploiters, these features of moral outrage would help mobilize coalitional behavior and reduce the risks and costs of punishments.<sup>1</sup> Given this, we propose that exploiter formidability influences moral outrage conditioned by the seriousness

---

<sup>1</sup> Another argument in relation to how humans could reduce the costs of aggressing against formidable individuals and at the same time hinder further exploitation is offered by Homicide Adaptation Theory (Buss, 2005; Duntley and Buss, 2005). According to Homicide Adaptation Theory, the human mind contains adaptations designed to effectively kill others when necessary, and to do so while minimizing the risks involved. For example, the activation of these adaptations could facilitate a search for circumstances under which the victim would not have time to react (e.g., an ambush at night). We would not expect any such adaptations to become activated, however, unless the initial exploitative act had vital survival implications (e.g., mate poaching, lethal aggression, massive theft, status challenges). Mobilization of coalitional behavior, in contrast, should operate at significantly lower levels of exploitation severity (e.g., small-scale theft, cheating, breaking social rules).

of the exploitive act. For less serious acts, the risks of aggressing against formidable individuals outweigh future threats and, hence, such exploitive acts are more acceptable when performed by a formidable individual than by a non-formidable individual. For more serious acts, the cost-benefit ratio at some point shifts and more moral outrage is produced against the formidable individual; not necessarily to motivate the individual to engage in punitive behavior on his own, but to mobilize coalitional aggression against the formidable exploiter.

While experimental studies show that victims of exploitive acts (e.g., non-reciprocation, free-riding) in particular are angered and willing to punish exploiters, studies document that unaffected third parties also feel angered and are willing to sacrifice personal winnings in order to punish exploiters (Carpenter and Matthews, 2005; Fehr and Fischbacher, 2004; Fehr and Gächter, 2000, 2002; O’Gorman, Wilson, and Miller, 2005; Ostrom, Walker, and Gardner, 1992). Also, norm enforcement through third party sanctioning is common in most small-scale hunter-gatherer societies (Cronk, Chagnon, and Irons, 2000; Sober and Wilson, 1998). Third parties mobilized to engage in the physical punishment of exploiters will often face the same cost-benefit-ratios as second parties. For example, direct physical aggression against a formidable exploiter is equally costly for a second or a third party. Furthermore, a formidable exploiter may also pose a problem for individuals who were not direct victims of the exploitive act – either because of involvement with the victim or because they could themselves be the next victim (for discussion, see Petersen et al., 2010). In this way, we expect the above selection pressures to not only apply to second parties but also shape how third parties factor formidability into their reactions to exploitation.

Studies in criminology have already provided some evidence for the proposed argument, that exploiter formidability influences third party moral outrage conditioned by the seriousness of the exploitive act. An experimental study with simulated civil trials found that male baby-faced defendants in crime cases involving negligent actions (i.e., in our perspective, less serious exploitive acts) were more likely to be convicted and, if convicted, were punished more compared to male mature-faced defendants (Berry and Zebrowitz-McArthur, 1988). In crime cases involving intentional actions (i.e., more serious exploitive acts), male baby-faced defendants were, in contrast, less likely than mature-faced male defendants to be convicted. A field study of 506 cases in small claims court in USA somewhat corroborated these results (Zebrowitz and McDonald, 1991). Similarly, other judicial studies suggest that high status individuals are punished less for minor transgressions, but conversely punished more for large transgressions (Levine, 1989; Rosoff, 1989; Wiggins, Dill, and Schwartz, 1965).

To corroborate the evolutionary interpretation of these effects, additional predictions can be offered. For second as well as third parties, an exploiter’s formidability is a factor of particular importance when punishment is physical (or, at least, when one runs the risk of direct physical confrontation). Cross-cultural evidence suggests that, over evolutionary time, males have engaged much more than females in both coalitional and individual aggression, and that the aggression has predominately been directed against other males (e.g., Chagnon, 1983; Daly and Wilson, 1988; Van Vugt, De Cremer, and Janssen, 2007; Wrangham and Peterson, 1996; see also Tooby and Cosmides, 2010). Therefore, the predictions specified above only apply in a straightforward manner to males. It is males that should fear the formidable male exploiter and be responsible for mobilizing against him. In the case of females, different possibilities suggest themselves from an evolutionary perspective. First, females could be expected to defer to formidable males across all situations such that they would express less outrage against formidable exploiters in both serious and non-serious

instances of exploitation. Second, the fact that men on average have 90 percent more upper body strength than women (Bohannon, 1997; Stoll, Huber, Seifert, Michel, and Stucki, 2000) could act as an inhibiting force on female reactions against *all* male exploiters (whether formidable or non-formidable relative to other males). If this is true, females should express less outrage than males against male exploiters. Finally, given that female third party punishment would only rarely take the form of direct confrontation and rather involve the imposition of indirect social costs, e.g., in the form of gossiping (see Fry, 1998), it is plausible that females simply do not modulate their outrage to the formidability of the exploiter.

## **Materials and Methods**

### *Participants*

To test these possibilities, a sample of 1621 Danish respondents (812 men) in age from 18 to 70 ( $M = 43.91$ ,  $SD = 13.91$ ) were recruited via the survey-company Yougov Zapera. To ensure national representativeness, the sample was based on quota sampling and drawn to match the national distribution on gender, sex and geographical region. In addition, the data is weighted on these parameters prior to analysis. The experiment was conducted online in November 2009.

### *Experimental design*

The study used a 2 (formidable vs. non-formidable exploiter) x 2 (serious vs. trivial exploitation) between-subjects experimental design. Subjects were randomly allocated to one of the four conditions. To manipulate formidability, we relied on manipulated head-shot pictures of the exploiter. To manipulate the seriousness of the exploitive act, two fictional scenarios were created – each with an exploitive act being committed. To ensure that the participants paid close attention to the pictured exploiter, participants were first presented with the picture (either the formidable or the non-formidable version) and were asked to rate the man for personality traits. On the second slide, the participants were presented with the picture and the exploitation scenario (either the serious or the trivial exploitation scenario). On the third slide, subjects were presented with the picture and the items measuring moral outrage.

### *Formidability*

To manipulate formidability, we rely on facial features. As testosterone is both the main determinant of physical strength (lean muscle mass) (Griggs et al., 1989; Lassek and Gaulin, 2009; Page et al., 2005) and of sexual dimorphism of the face (Penton-Voak and Chen, 2004; Pound, Penton-Voak, and SurrIDGE, 2009; Weston, Friday, and Lio, 2007), it has been found that facial masculinity is an indicator of actual physical strength (Fink, Neave, and Seydel, 2007; Sell et al., 2009) as well as aggressive and exploitive behavior in men both in real life and in laboratory settings (Carré and McCormick, 2008; Mazur and Booth, 1998; Stirrat and Perrett, 2010). Studies also show that humans consistently associate facial masculinity with aggressive and exploitive behavior (Carré et al., 2009; Perrett et al., 1998), while facial immaturity (baby-faced: large eyes, high eye-brows, small chin) is associated with perceived submissiveness, physical weakness, incompetence, but also honesty and kindness (Keating, 1985; McArthur and Apatow, 1983; McArthur and Berry, 1987). Facial

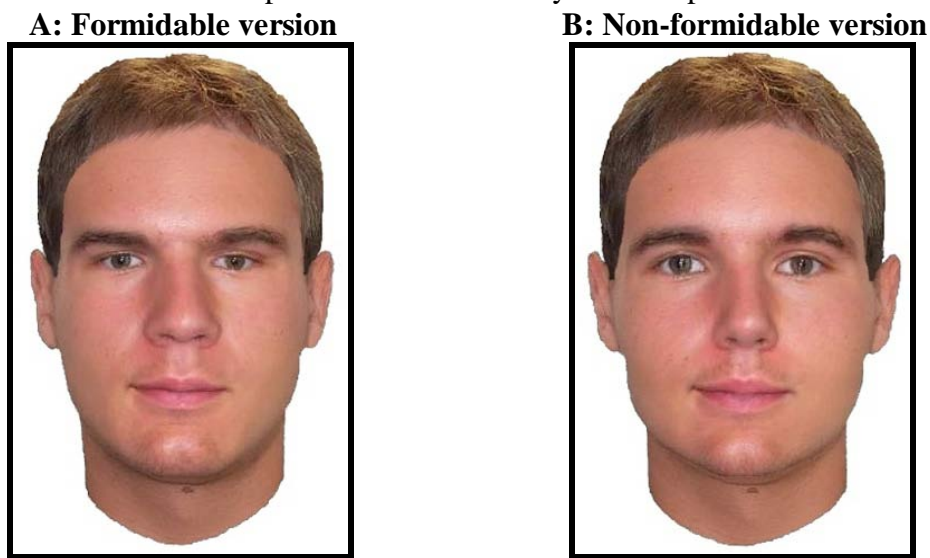
masculinity is related to traits such as a strong jaw, broad cheekbones, lowered eyebrows, and a large facial width-to-height ratio (Weston et al., 2007).

Specifically, we digitally constructed two versions of a head-shot photo of a young man; the two versions only differed in facial masculinity. The photo was selected from the Center for Vital Longevity Face Database (Minear and Park, 2004). Manipulation of facial masculinity was made in Facegen Modeller®, which is a computer program enabling manipulations of distinct and selectable facial characteristics using advanced vector calculations (Facegen Modeller version 3.4; see also Oosterhof and Todorov, 2008, for illustrative material). To increase realism, the two versions of masculinity were, respectively, morphed with the original photo using Magicmorph®.

In addition to the formidability-related trait of dominance (cf. Sell et al., 2009), recent studies in face perception suggest that another main trait spontaneously attributed to human faces is trustworthiness (Oosterhof and Todorov, 2008). To validate our experimental conditions on these two dimensions, we had 20 subjects rate the faces for dominance and trustworthiness. Subjects were presented with one of the faces and asked whether they agreed or disagreed that the terms "dominant" and "untrustworthy," respectively, applied to the person. Answers were obtained on seven-point scales with "Completely Agree" scored as 1 and "Completely Disagree" scored as 7. As expected, the formidable face was rated as significantly more dominant than the non-formidable face ( $M_{\text{formidable}} = 2.27$ ,  $M_{\text{non-formidable}} = 3.44$ ,  $t(18) = 2.24$ ,  $p = .038$ , two-tailed), while there was no significant difference in perceptions of untrustworthiness ( $M_{\text{formidable}} = 4.18$ ,  $M_{\text{non-formidable}} = 3.79$ ,  $t(18) = -.53$ ,  $p = .60$ , two-tailed).

In the actual study, we added hair to the two faces to avoid skinhead-stereotyping (see Bodenhausen and Wyer, 1985). The resulting formidable (figure 1A) and non-formidable (figure 1B) versions of the subject were presented in JPEG format in a 350 x 510 resolution on a white background.

**Figure 1.** Pictures used to manipulate the formidability of the exploiter



For additional validation and control, the subjects in the actual study rated the extent to which they perceived the man in the picture to be *dominant*, *physically strong*, *physically*

*attractive, friendly, trustworthy, extrovert, responsible, laid-back, and open-minded.* All nine items were rated on a five-point scale from 1: not at all to 5: extremely. These ratings were performed when subjects were first presented with the picture prior to receiving any information about the pictured person's exploitative tendencies. Using factor analysis on the nine traits on which the faces were rated revealed a clear two-dimensional structure (see Table 1).

**Table 1.** Factor analysis of traits ratings

<b>Trait</b>	<b>Factor 1: Trustworthiness</b>	<b>Factor 2: Formidability</b>
Dominant	-.18	<b>.65</b>
Physically strong	.30	<b>.60</b>
Friendly	<b>.81</b>	-.26
Trustworthy	<b>.80</b>	-.07
Open-minded	<b>.74</b>	-.12
Extrovert	<b>.65</b>	.04
Responsible	<b>.65</b>	.12
Physically attractive	<b>.59</b>	.20
Laid-back	<b>.24</b>	.01
Eigen-value	3.68	1.46

*Notes:* Principal Axis Factoring was used as extraction method and Oblimin was used as rotation method. Only the two displayed factors had eigen-values above 1.

As expected, the two formidability-related traits (dominant, physically strong) loaded on a separate factor, while the rest of the traits loaded much higher on a distinct single common factor. In line with previous research, trustworthiness and friendliness has the highest loadings on this other factor. The factor scores computed from the factor analysis for each subject was saved. Validating the facial manipulations, the formidable individual was judged as significantly higher than the non-formidable individual on the dominance dimension ( $M_{\text{formidable}} = .25$ ,  $M_{\text{non-formidable}} = -.25$ ,  $t(1619) = 12.76$ ;  $p < .001$ ). However, in contrast to the pilot results, the formidable individual was also judged as significantly less trustworthy ( $M_{\text{formidable}} = -.27$ ,  $M_{\text{non-formidable}} = .28$ ,  $t(1619) = -12.24$ ;  $p < .001$ ). Given some previous research (Carré et al., 2008; Stirrat and Perrett, 2010) this is not unexpected. This difference creates a potential confound as research shows that offender trustworthiness has a large influence on third party sanctioning (Baldwin and Kleinke, 1994; Feather and Atchison, 1998; Landy and Aronson, 1969; Weiten, 1980). As this difference between the faces is conceptually and empirically distinct from our main theoretical dimension of interest (i.e., formidability), we include the subjects' perception of the trustworthiness of the exploiter in the analyses as a control variable.

*Exploitation scenarios*

In order to evoke moral outrage in respondents, two fictional scenarios were created – each with an exploitative act being committed. Fictional scenarios have been used by others

investigating context effects on moralistic punishment (Lieberman and Linke, 2007; O’Gorman, Wilson, and Miller, 2005), and fictional scenarios have been found to be quite adept at invoking the same emotional reactions as observations in real-life (Berthoz, Armony, Blair, and Dolan, 2002; Wilson and O’Gorman, 2003). The exploitative acts were constructed with some degree of ambiguity that enabled subjects to react either strongly towards the exploitation or to downplay it. The trivial exploitative act involved a scenario where the individual was cutting in line, while arguing that he was in a hurry. The serious exploitative act involved a scenario where the individual destroyed private property for no apparent reason. The wordings of the scenarios appear in the appendix.

### *Moral outrage*

To measure the degree of third party moral outrage, eight items were constructed based on the theoretical framework from the moral-convention distinction task (see the appendix, table A1 and A2) (Kelly, Stich, Haley, Eng, and Fessler, 2007; Turiel, 1983). All items were measured on a five-point scale from 1: strongly disagree to 5: strongly agree. The items are essentially worded in the same way across the two scenarios with slight variations given the specifics of the different scenarios. To allow for comparisons despite these variations,  $z$ -scores for the moral outrage scales are computed independently for each scenario. Hence, a value of 0 on the scale corresponds to the average level of outrage in the given scenario and positive and negative numbers indicate numbers of standard deviations in positive or negative direction from this average. Two items were excluded from the final moral outrage scale, as these items correlated less than  $r = .20$  with the other six items (final  $\alpha = .83$ )

## **Results**

Given that we hold distinct expectations for the reactions of males and females, the two sexes are analyzed separately. In the case of males, our theoretical expectations are clear and imply that third party outrage is modulated by the two-way interaction between exploitation severity and exploiter formidability. More specifically, we expect that more outrage is expressed against formidable exploiters committing serious exploitative acts while less outrage is expressed against formidable exploiters committing trivial exploitive acts. Our initial analysis reveals a significant main effect of exploiter formidability on expressed outrage ( $M_{\text{non-formidable}} = -.07$ ,  $M_{\text{formidable}} = .08$ ,  $t(809) = 2.06$ ,  $p = .04$ , two-tailed). On average, across the two scenarios (i.e., the more and the less serious violations), more outrage is expressed against the formidable exploiter. To investigate our core claim, that this effect is modulated by the severity of the exploitive act, we test the effect of the two-way interaction between the formidability and severity manipulation. Contrary to our expectations, this interaction is insignificant ( $F_{1,808} = 2.11$ ,  $p = .147$ , two-tailed). To explore whether this result is confounded by the differences in trustworthiness observed between the face manipulations, we include the perceived trustworthiness of the exploiter as a covariate in the analysis. As shown in Table 2, this analysis reveals the existence of a significant three-way interaction effect between exploiter formidability, exploitation severity and the perceived trustworthiness of the exploiter ( $F_{1,804} = 7.37$ ,  $p = .007$ , two-tailed).

**Table 2.** Male moral outrage towards exploiters as modulated by seriousness of exploitation and characteristics of the exploiter

	Coefficient
Intercept	.04 (.07)
Formidability of exploiter	-.04 (.11)
Seriousness of exploitation	-.14 (.10)
Perceived trustworthiness of exploiter	-.09 (.08)
Formidability × Seriousness	.23 (.15)
Formidability × Trustworthiness	-.24* (.11)
Seriousness × Trustworthiness	-.17 (.13)
Formidability × Seriousness × Trustworthiness	.47** (.17)
<i>R</i> <sup>2</sup>	.05

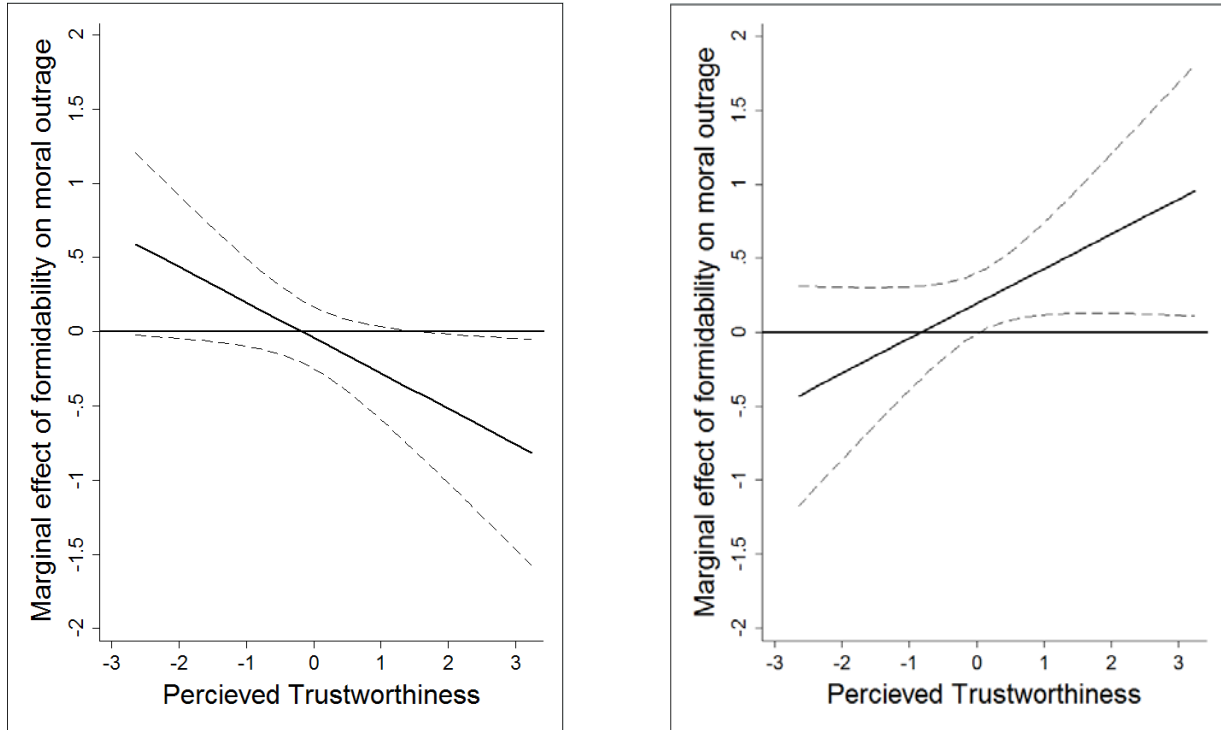
*Notes:* *N* = 812. Entries are OLS regression coefficients with standard errors in parentheses. The dependent variable, moral outrage, and “Perceived trustworthiness of exploiter” are z-scored variables with high values corresponding to high levels of outrage and perceived trustworthiness. “Formidability of exploiter” and “Seriousness of exploitation” are dichotomous experimental variables in a 2\*2 factorial design (0 = exploiter is non-formidable / exploitation is trivial, 1 = exploiter is formidable / exploitation is serious). \**p* < .05, \*\**p* < .01.

Based on Table 2, Figure 2A and B provides a graphical display of the three-way interaction effect. The figure is a marginal effect plot with associated 95 percent confidence intervals (see Kam and Franzese, 2007 for statistical procedures). It shows how the marginal effect of exploiter formidability changes as a function of perceived trustworthiness. A positive effect – when statistically different from zero (i.e., zero is not included in the confidence interval) – implies that more outrage is expressed towards the formidable exploiter relative to the non-formidable exploiter, whereas a negative effect – when statistically different from zero – implies less outrage expressed towards the formidable exploiter. Figure 2A displays the marginal effect plot for the trivial exploitive act, while Figure 2B displays the marginal effect for the serious exploitive act.

**Figure 2.** Marginal effect of exploiter formidability on male moral outrage in the face of trivial (A) and serious (B) exploitation as a function of the exploiter’s perceived trustworthiness. Includes 95 percent confidence interval for the difference.

**A. Trivial Exploitation**

**B. Serious Exploitation**



*Notes:* The marginal effects are calculated based on Table 2 and correspond to the coefficient of the formidability variable in Table 2 given the values of the other two variables in the three-way interaction term. A positive effect indicates that the formidable exploiter triggers *more* outrage than the non-formidable exploiter. The confidence intervals indicate when this difference is significant at the .05 level. This is the case when 0 (i.e., the value indicating no effect of the experimental formidability manipulation) is not included in the confidence interval (see Kam and Franzese, 2007, for statistical procedures).

As can be seen, the predicted modulation of outrage as a function of exploiter formidability and exploitation severity emerges when the exploiter is initially deemed to be trustworthy. For the trivial exploitive act (Figure 2A), subjects begin to discriminate between the formidable and non-formidable exploiter as the perceived trustworthiness of the exploiter increases. Specifically, males become significantly *less* outraged at formidable exploiters relative to non-formidable exploiters. For the serious exploitive act, the interaction effect is as expected in the opposite direction. As shown in Figure 2B, when exploitation is serious and the exploiter is perceived as trustworthy, males become *more* outraged when the exploiter is formidable. These results show that while males experience more outrage against formidable individuals when they commit serious acts of exploitation, trivial exploitive acts from formidable individuals activate less outrage than trivial acts from non-formidable others, at least when these individuals are perceived as initially trustworthy. Under this circumstance, the theoretical predictions are supported for males.

The analyses also reveal how exploiter formidability can shape third party outrage among males in another way. In general, there is a negative effect of the perceived trustworthiness on outrage, such that more outrage is expressed towards those who were initially deemed to be untrustworthy (zero-order effect:  $r = -.17$ ,  $p < .0001$ , two-tailed). Given the effect of the formidability manipulation on perceived trustworthiness (cf. the Materials and Methods section), trustworthiness provides an indirect path through which formidability, *ceteris paribus*, increases the outrage expressed towards an exploiter. In line with this, the initially reported main effect of the formidability manipulation turns insignificant ( $F_{1,808} = .79$ ,  $p = .37$ , two-tailed) when controlled for perceived trustworthiness. As the relationships reported in Figure 2A and B do include perceived trustworthiness, any such indirect effects exist in addition to the predicted modulation of outrage as a function of both exploiter formidability and exploitation seriousness.

Given the male-bias for coalitional aggression, we do not expect females to express more outrage (i.e., display a mobilization effect) against formidable exploiters. However, it is less clear whether they, on a priori grounds, will treat formidable exploiters more leniently across scenarios. To explore the different possibilities, we first try to replicate the patterns of effects among males. As expected, both the two-way interaction – between exploiter formidability and exploitation seriousness ( $F_{1,805} = 1.32$ ,  $p = .25$ , two-tailed) – and the three-way interaction – between exploiter formidability, exploitation seriousness, and perceived trustworthiness of the exploiter ( $F_{1,801} = .25$ ,  $p = .61$ , two-tailed) – are insignificant. For females, the effect of exploiter formidability does not, in other words, depend on the seriousness of the exploitive act. To test whether females treat all formidable exploiters more leniently, independent of exploitation seriousness, we next investigate the existence of a main effect of exploiter formidability on female outrage. Again, we find no such effect ( $M_{\text{non-formidable}} = -.06$ ,  $M_{\text{formidable}} = .04$ ,  $t(808) = 1.39$ ,  $p = .16$ , two-tailed).

In tandem, these results show that third party outrage among females is not modulated by differences in formidability between male exploiters. One simple explanation for this result could be that females are less attentive to male differences in formidability. However, previous research has found no such sex differences (Sell et al., 2009), and both theory and data suggest that male differences in physical strength is an important parameter in mate selection processes for females (Buss and Schmitt, 1993; Dixson, Halliwell, East, Wignarajah, and Anderson, 2003; Frederick and Haselton, 2007; Little et al., 2010; Maisey, Vale, Cornelissen, and Tovée, 1999). At the outset, it is therefore unlikely that females do not pay attention to differences in formidability. In line with this, our data reveals no sex differences in the effect of our experimental manipulation of exploiter formidability on the ratings of perceived formidability. Specifically, we find an insignificant two-way interaction between the sex of the subject and the formidability manipulation on the subjects' perception of the formidability of the exploiter ( $F_{1,1617} = .22$ ,  $p = .64$ , two-tailed). Hence, it is not that females do not attend to male differences in formidability, they just don't modulate their outrage to it. As argued in the theory section, this could reflect the fact that (1) males on average are 90 percent stronger than females, which would imply that even a non-formidable male would be stronger than many females, or (2) females do not reason about third party situations in terms of direct physical confrontations where differences in formidability would be important. The first explanation entails that the lack of effect of male differences in formidability for female outrage reflects that females defer to all male exploiters. This explanation thereby requires that females in general express less outrage than males towards

male exploiters. Importantly, this is not the case. There are no sex differences in the outrage individuals express towards the exploiters ( $M_{\text{males}} = .01$ ,  $M_{\text{females}} = -.02$ ,  $t(1620) = .49$ ,  $p = .63$ , two-tailed). This leaves us with the second explanation: viz., females do not reason about third party involvement in terms of physical confrontation, presumably because such confrontation has been a male domain over evolutionary history.

## **Discussion**

This study adds to an increasing number of studies demonstrating that third party moral responses to social violations are modulated by the characteristics of the exploiter. Consistent with these studies, the above findings demonstrate that the trustworthiness of the exploiter is a major component in this regard. Individuals, whose personalities are gauged to be geared towards continued exploitation, elicit large levels of outrage. Considering the fact that moral outrage serve as output of systems, in part, designed to deter individuals from transgressing again (Petersen et al., 2010; Price et al., 2002; Sunstein, 2005), the evolutionary importance of exploiter trustworthiness is obvious.

At the same time, the present study goes significantly beyond previous studies. By using very subtle face-manipulations in a nationally representative sample, we have provided evidence that exploiters' formidability is an additional determinant of moral outrage in third parties. When exploiters were perceived as trustworthy, we found that males are less outraged when confronted with formidable offenders having committed a trivial rule violation which is in line with traditional models of animal behavior in conflict situations. Yet, supporting alternative theories about coalitional control of powerful individuals, situations involving more serious transgressions activated heightened outrage among males against formidable exploiters. In other words, if a formidable individual keeps exploitation within limits, he is excused. But if a formidable male reveals massive exploitive tendencies, other males mobilize against him. We also replicated a finding in previous studies that perceptions of trustworthiness are negatively influenced by formidability (Carré et al., 2009; Stirrat and Perrett, 2010). For males, trustworthiness thus serves as an additional channel through which the formidability of exploiters influences moral outrage.

In the case of females, results showed that females did not modulate their expressed outrage in response to exploiters' formidability. The relevance of exploiter formidability for third parties arises, in particular, when engaging in direct physical punishment of the exploiter. Given this, the lack of effect among females is expected and, as indicated in previous research, direct aggression has been a male domain over evolutionary history (e.g., Chagnon, 1983; Van Vugt, De Cremer, and Janssen, 2007; Wrangham and Peterson, 1996; see also Tooby and Cosmides, 2010). As a consequence, we suggest, females do not reason about third party sanctioning in terms of physical confrontation with the exploiter. While the results provide no direct evidence of this, analyses indicate that the lack of formidability-effect on female outrage cannot be explained by alternative interpretations, such as the possibility that females are inattentive to male differences in formidability, or that females defer (i.e., express low levels of outrage) to all male exploiters.

Returning to our main hypothesis, the predicted effects of exploiter formidability on male third party outrage only appeared when subjects initially perceived the exploiters as trustworthy. In terms of selection pressures, the effect of exploiter formidability on moral outrage was predicted based on the fact that formidability would influence the costs

associated with future exploitation. In the case of massively untrustworthy individuals such as psychopaths, differences in formidability might be less predictive of future costs as even non-formidable individuals can inflict massive costs if their intentions are malicious enough. The present results could reflect this. In that perspective, a counter-exploitation psychology geared towards estimating the costs of future exploitation would be designed to disregard other person attributes when confronted with highly untrustworthy individuals.

These considerations also point to directions for future research. First of all, it would be important to replicate the results using manipulations that independently manipulated the trustworthiness and formidability of exploiters. In this way, it would be possible to decide whether the present effects of trustworthiness ratings reflect the operations of mechanisms designed to react differently to trustworthy or untrustworthy exploiters or, whether they result from some individual differences among high and low trusters (e.g., Yamagishi, 2003). Second, it would be important to replicate the results using scenarios that more equally match antecedent circumstances. For example, in the present case, one scenario reflects a criminal offense, the other does not. Correspondingly, it would be interesting to investigate whether the results might be generalized to other moral domains (i.e., violations in the domains of purity/sanctity, ingroup/loyalty, or authority/respect) (Haidt, 2007). Finally, in future studies, it would be important to provide additional and independent evidence that the function of moral outrage in the face of a serious formidable exploiter is to coordinate joint coalitional attention. Are people, for example, more inclined to convey information about such exploiters to others? Are they easier to remember? Do women also express outrage towards formidable offenders when given coalitional cues? Affirmative answers to such questions could provide relevant evidence of coalitional mobilization.

All in all, even though moral outrage is designed to fuel moralistic punishment against *all* social exploiters, this study provides additional support for human outrage being a fine-tuned strategic resource deployed according to subtle cues. Given the involved costs and benefits, male moral outrage is withheld against formidable individuals if they do not excessively exploit their bargaining power. Yet, if that line is crossed, the formidable individual faces severe outrage among male third parties. As the existence of this counter-exploitation circuitry would in itself work as a selective force, we might expect dominant individuals to intuitively recognize this set of dynamics.

**Received 28 September 2010; Revision submitted 14 February 2011; Accepted 2 March 2011**

## **References**

- Ahola, A.S., Christianson, S.Å., and Hellström, Å. (2009). Justice needs a blindfold: Effects of gender and attractiveness on prison sentences and attributions of personal characteristics in a judicial process. *Psychiatry, Psychology and Law*, 16, 90-100.
- Archer, J. (1988). *The behavioural biology of aggression*. New York: Cambridge University Press.
- Archer, J., and Benson, D. (2008). Physical aggression as a function of perceived fighting ability and provocation: An experimental investigation. *Aggressive Behavior*, 34, 9-24.

- Archer, J., and Thanzami, V. (2007). The relation between physical aggression, size and strength, among a sample of young Indian men. *Personality and Individual Differences*, 43, 627-633.
- Archer, J., and Thanzami, V. (2009). The relation between mate value, entitlement, physical aggression, size and strength among a sample of young Indian men. *Evolution and Human Behavior*, 30, 315-321.
- Baldwin, M.R., and Kleinke, C.L. (1994). Effects of severity of accident, intent, and “alcoholism is a disease” excuse on judgments of a drunk driver. *Journal of Applied Social Psychology*, 24, 2097-2109.
- Berry, D.S., and Zebrowitz-McArthur, L. (1988). What’s in a face?: Facial maturity and the attribution of legal responsibility. *Personality and Social Psychology Bulletin*, 14, 23-33.
- Berthoz, S., Armony, J.L., Blair, R.J.R., and Dolan, R.J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain: A Journal of Neurology*, 125, 1696-1708.
- Bodenhausen, G.V., and Wyer, R.S. (1985). Effects of stereotypes in decision making and information-processing strategies. *Journal of personality and social psychology*, 48, 267-282.
- Boehm, C. (1993). Egalitarian behavior and reverse dominance hierarchy. *Current Anthropology*, 34, 227-254.
- Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Harvard University Press.
- Boehm, C. (2000). Conflict and the evolution of social control. *Journal of Consciousness Studies*, 7, 79.
- Bohannon, R.W. (1997). Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. *Archives of Physical Medicine and Rehabilitation*, 78, 26–32.
- Buss, D.M. (2005). *The murderer next door: Why the mind is designed to kill*. New York: Penguin Press.
- Buss, D.M., and Schmitt, D.P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review*, 100, 204-232.
- Carpenter, J.P., and Matthews, P.H. (2005). Norm enforcement: Anger, indignation or reciprocity? Unpublished manuscript.
- Carré, J.M., and McCormick, C.M. (2008). In your face: Facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society B: Biological Sciences*, 275, 2651-2656.
- Carré, J.M., McCormick, C.M., and Mondloch, C.J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, 20, 1194-1198.
- Cashdan, E. (1980). Egalitarianism among hunters and gatherers. *American Anthropologist*, 82, 116-120.
- Chagnon, N. (1983). *The Yanomamö*. New York: Holt, Rinehart and Winston.
- Cronk, L., Chagnon, N., and Irons, W. (2000). *Human behavior: An anthropological perspective*. New York: Aldine De Gruyter.
- Cummins, D. (1999). Cheater detection is modified by social rank: The impact of dominance on the evolution of cognitive functions. *Evolution and Human Behavior*, 20, 229-248.

- Cummins, D. (2005). Dominance, status, and social hierarchies. In D.M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 676-697). Hoboken: John Wiley and Sons Inc.
- Daly, M., and Wilson, M. (1988). *Homicide*. Hawthorne, NY: Aldine.
- Darley, J.M., and Pittman, T.S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, 7, 324-336.
- Diamond, J. (1997). *Guns, germs, and steel*. New York: W.W. Norton.
- Dixson, A.F., Halliwell, G., East, R., Wignarajah, P., and Anderson, M. (2003). Masculine somatotype and hirsuteness as determinants of sexual attractiveness to women. *Archives of Sexual Behavior*, 32, 29-39.
- Dugatkin, L.A. (1997). *Cooperation among animals: An evolutionary perspective*. Oxford: Oxford University Press.
- Duntley, J.D., and Buss, D.M. (2005). The plausibility of adaptations for homicide. In P.S.L. Caruthers and S. Stich (Eds.), *The innate mind: Structure and contents* (pp. 291-304). New York: Oxford University Press.
- Feather, N.T., and Atchison, L. (1998). Reactions to an offence in relation to the status and perceived moral character of the offender. *Australian Journal of Psychology*, 50, 119-127.
- Fehr, E., and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25, 63-87.
- Fehr, E., and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90, 980-994.
- Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140.
- Felson, R.B. (1996). Big people hit little people: Sex differences in physical power and interpersonal violence. *Criminology*, 34, 433-452.
- Fink, B., Neave, N., and Seydel, H. (2007). Male facial appearance signals physical strength to women. *American Journal of Human Biology*, 19, 82-87.
- Frederick, D.A., and Haselton, M. (2007). Why is muscularity sexy? Tests of the fitness indicator hypothesis. *Personality and Social Psychology Bulletin*, 33, 1167-1183.
- Fry, D.P. (1998). Anthropological perspectives on aggression: Sex differences and cultural variation. *Aggressive Behavior*, 24, 81-95.
- Furer-Haimendorf, C. (1967). *Morals and merit: A study of values and social controls in South Asian societies*. Chicago: University of Chicago Press.
- Gallup, A.C., White, D.D., and Gallup Jr, G.G. (2007). Handgrip strength predicts sexual behavior, body morphology, and aggression in male college students. *Evolution and Human Behavior*, 28, 423-429.
- Gardner, P. (1991). Foragers' pursuit of individual autonomy. *Current Anthropology*, 32, 543-558.
- Greene, J. and Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6, 517-523.
- Griggs, R.C., Kingston, W., Jozefowicz, R.F., Herr, B.E., Forbes, G., and Halliday, D. (1989). Effect of testosterone on muscle mass and muscle protein synthesis. *Journal of Applied Physiology*, 66, 498-503.
- Haidt, J. (2003). The moral emotions. In R.J. Davidson, K.R. Scherer, and H.H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852-870). Oxford: Oxford University Press.

- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998-1002.
- Hammerstein, P., and Parker, G.A. (1982). The asymmetric war of attrition. *Journal of Theoretical Biology*, 96, 647-682.
- Huntingford, F.A., and Turner, A.K. (1987). *Animal conflict*. New York: Chapman and Hall.
- Kahneman, D., and Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). New York: Cambridge University Press.
- Kam, C.D., and Franzese, R.J. (2007). *Modeling and interpreting interactive hypotheses in regression analysis*. Michigan: University of Michigan Press.
- Keating, C.F. (1985). Gender and the physiognomy of dominance and attractiveness. *Social Psychology Quarterly*, 48, 61-70.
- Kelly, D., Stich, S., Haley, K.J., Eng, S.J., and Fessler, D.M.T. (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language*, 22, 117-131.
- Landy, D., and Aronson, E. (1969). The influence of the character of the criminal and his victim on the decisions of simulated jurors. *Journal of Experimental Social Psychology*, 5, 141-152.
- Lassek, W.D., and Gaulin, S.J.C. (2009). Costs and benefits of fat-free muscle mass in men: Relationship to mating success, dietary requirements, and native immunity. *Evolution and Human Behavior*, 30, 322-328.
- Levine, J.M. (1989). Reaction to opinion deviance in small groups. In P.B. Paulus (Ed.), *Psychology of group influence (2nd ed.)* (pp. 187-231). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Lieberman, D., and Linke, L. (2007). The effect of social category on third party punishment. *Evolutionary Psychology*, 5, 289-305.
- Little, A.C., Saxton, T.K., Roberts, S.C., Jones, B.C., DeBruine, L.M., Vukovic, J., Perrett, D.I., et al. (2010). Women's preferences for masculinity in male faces are highest during reproductive age range and lower around puberty and post-menopause. *Psychoneuroendocrinology*, 35, 912-920.
- Maisey, D.S., Vale, E.L.E., Cornelissen, P.L., and Tovée, M.J. (1999). Characteristics of male attractiveness for women. *Lancet*, 353, 1500.
- Maynard Smith, J., and Price, G.R. (1973). The logic of animal conflict. *Nature*, 246, 15-18.
- Mazur, A., and Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences*, 21, 353-397.
- McArthur, L.Z., and Apatow, K. (1983). Impressions of baby-faced adults. *Social Cognition*, 2, 315-342.
- McArthur, L.Z., and Berry, D.S. (1987). Cross-cultural agreement in perceptions of babyfaced adults. *Journal of Cross-Cultural Psychology*, 18, 165-192.
- Miner, M., and Park, D. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments & Computers, Special Issue: Web-based archive of norms, stimuli, and data: Part 2*, 36, 630-633.
- O'Gorman, R., Wilson, D.S., and Miller, R.R. (2005). Altruistic punishing and helping differ in sensitivity to relatedness, friendship, and future interactions. *Evolution and Human Behavior*, 26, 375-387.

- Oosterhof, N.N., and Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 11087-11092.
- O’Gorman, R., Wilson, D.S., and Miller, R.B. (2005). Altruistic punishing and helping differ in sensitivity to relatedness, friendship, and future interactions. *Evolution and Human Behavior*, *26*, 375-387.
- Ostrom, E., Walker, J., and Gardner, R. (1992) Covenants with and without a sword: Self-governance is possible. *The American Political Science Review*, *86*, 404-417.
- Overbeck, J.R., Tiedens, L.Z., and Brion, S. (2006). The powerful want to, the powerless have to: Perceived constraint moderates causal attributions. *European Journal of Social Psychology*, *36*, 479-496.
- Page, S.T., Amory, J.K., Bowman, F.D., Anawalt, B.D., Matsumoto, A.M., Bremner, W.J., and Tenover, J.L. (2005). Exogenous testosterone (T) alone or with finasteride increases physical performance, grip strength, and lean body mass in older men with low serum T. *Journal of Clinical Endocrinology Metabolism*, *90*, 1502-1510.
- Parker, G.A. (1974). Assessment strategy and the evolution of fighting behaviour. *Journal of Theoretical Biology*, *47*, 223-243.
- Pellegrini, A.D., Roseth, C.J., Mliner, S., Bohn, C.M., Van Ryzin, M., Vance, N., Cheatham, C.L., and Tarullo, A. (2007). Social dominance in preschool classrooms. *Journal of Comparative Psychology*, *121*, 54-64.
- Penton-Voak, I.S., and Chen, J.Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior*, *25*, 229-241.
- Perrett, D.I., Lee, K.J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D.M., Henzi, S.P., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, *394*, 884-887.
- Petersen, M.B., Sell, S., Tooby, J., and Cosmides, L. (2010). Evolutionary psychology and criminal justice: A recalibrational theory of punishment and reconciliation. In H. Høgh-Olesen (Ed.), *Human morality & sociality: Evolutionary & comparative perspectives* (pp. 72-131). London: Palgrave Macmillan.
- Pound, N., Penton-Voak, I.S., and SurrIDGE, A. (2009). Testosterone responses to competition in men are related to facial masculinity. *Proceedings of the Royal Society B*, *276*, 153-159.
- Price, M.E., Cosmides, L., and Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior*, *23*, 203-231.
- Rosoff, S.M. (1989). Physicians as criminal defendants: Specialty, sanctions, and status liability. *Law and Human Behavior*, *13*, 231-236.
- Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., and Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society B*, *276*, 575-584.
- Sell, A., Tooby, J., and Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 15073-15078.
- Sell, A., Bryant, G.A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., Krauss, A., and Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B*. Published online before print, June 16.

- Sober, E., and Wilson, D.S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.
- Stewart, J.E. (1980). Defendant's attractiveness as a factor in the outcome of criminal trials: An observational study. *Journal of Applied Social Psychology, 10*, 348-361.
- Stirrat, M., and Perrett, D.I. (2010). Valid facial cues to cooperation and trust. *Psychological Science, 21*, 349-354.
- Stoll, T., Huber, E., Seifert, B., Michel, B.A., and Stucki, G. (2000). Maximum isometric muscle strength: Normative values and gender-specific relation to age. *Clinical Rheumatology, 19*, 105-111.
- Sunstein, C.R. (2005). Moral heuristics. *Behavioral and Brain Sciences, 28*, 531-573.
- Tetlock, P.E., Kristel, O.V., Elson, S.B., Green, M.C., and Lerner, J.S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*, 853-870.
- Tooby, J., and Cosmides, L. (2010). Groups in mind: The coalitional roots of war. In H. Høgh-Olesen, (Ed.), *Human morality & sociality: Evolutionary & comparative perspectives* (pp. 191-234). London: Palgrave Macmillan.
- Tooby, J., Cosmides, L., and Price, M. (2006). Cognitive adaptations for n-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics, 27*, 103-129.
- Tremblay, R.E., Schaal, B., Boulerice, B., Arseneault, L., Soussignan, R.G., Paquette, D., and Laurent, D. (1998). Testosterone, physical aggression, dominance, and physical development in early adolescence. *International Journal of Behavioral Development, 22*, 753-777.
- Turiel, E. (1983). *The development of social knowledge*. Cambridge: Cambridge University Press.
- Van Vugt, M., De Cremer, D., and Janssen, D.P. (2007). Gender differences in cooperation and competition: The male-warrior hypothesis. *Psychological Science, 18*, 19-23.
- Weiten, W. (1980). The attraction-leniency effect in jury research: An examination of external validity. *Journal of Applied Social Psychology, 10*, 340-347.
- Weston, E.M., Friday, A.E., and Lio, P. (2007). Biometric evidence that sexual selection has shaped the Hominin face. *PLoS ONE, 2*, e710.
- Wiggins, J.A., Dill, F., and Schwartz, R.D. (1965). On "status-liability". *Sociometry, 28*, 197-209.
- Wilson, D.S., and O'Gorman, R. (2003). Emotions and actions associated with norm-breaking events. *Human Nature, 14*, 277-304.
- Wrangham, R.W., and Peterson, D. (1996). *Demonic males: Apes and the origins of human violence*. Boston, MA: Houghton, Mifflin and Company.
- Yamagishi, T. (2003). Trust as a form of social intelligence. In K.S. Cook (Ed.), *Trust in society* (pp. 121-147). New York: Russell Sage Foundation.
- Zebrowitz, L.A., and McDonald, S.M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior, 15*, 603-623.

## **Appendix**

*Scenarios.* The trivial exploitation scenario was worded: “Imagine that the man in the picture is standing in a queue to buy a ticket for the train. A young girl stands in front of him in the queue and is talking on the phone. She is engrossed in the conversation, so for a moment, she is not aware that the queue has moved forward. The man therefore steps in front of her in the queue while saying that he is in a hurry.” The serious exploitation scenario was worded: “Imagine that the man in the picture has been arrested for smashing a shop window during a night out. He has been drinking heavily and can hardly remember a thing.”

**Table A1.** Moral outrage (scenario A: trivial exploitation)

1	The man’s behavior is totally unacceptable.
2	If the man hadn’t been in such a hurry, he probably wouldn’t have stepped in front of the women in the queue.
3	The man obviously doesn’t have any respect for others.
4	The man didn’t really do anything wrong.
5	People all over the world would be offended over the man’s behavior.
6	Regardless if the man broke any rules or not, one just cannot behave like that.
7	A lot of people would probably just shrug over the man’s behavior.
8	If the train station doesn’t have any formal rules on queuing culture, then one cannot blame the man for skipping ahead in the queue.

**Table A2.** Moral outrage (scenario B: serious exploitation)

1	The man’s crime is very serious.
2	If the man hadn’t been so drunk, he probably wouldn’t have smashed the window.
3	The man obviously doesn’t have any respect for others and their property.
4	The man’s action isn’t really that bad.
5	People all over the world would be furious with the man’s action.
6	No matter what the law says, I think that he should be thrown in prison for that kind of behavior.
7	A lot of people would probably just shrug over the man’s action.
8	If the police decide that he should get away with a warning, that’ll be alright.