The true trigger of shame: social devaluation is sufficient, wrongdoing is unnecessary

Theresa E. Robertson[a,*,1], Daniel Sznycer[b,**,1], Andrew W. Delton[a], John Tooby[c], Leda Cosmides[c]

[a] Stony Brook University, Stony Brook, NY, United States
[b] University of Montreal, Montreal, QC, Canada
[c] University of California, Santa Barbara, CA, United States

ARTICLE INFO

ABSTRACT

What is the trigger of shame? The information threat theory holds that shame is an evolved adaptation that is designed to limit the likelihood and costs of others forming negative beliefs about the self. By contrast, attributional theories posit that concerns over others' evaluations are irrelevant to shame. Instead, shame is triggered when a person attributes a negative outcome to their *self*, rather than to a particular act or circumstance. We conduct a strong test of the information threat hypothesis. In Study 1, participants imagined taking an action that, though morally unimpeachable, could be interpreted unfavorably by others. As predicted by the information threat theory, shame increased with the publicity of this act. In Study 2, participants played a public good game and then learned that the other participants either chose to keep interacting with them (inclusion) or not (exclusion)—ostensibly because of their contributions, but in fact randomly determined by the experimenter. Exclusion increased shame. Under-contribution did not. In fact, even the highest contributors tended to feel shame when excluded. These findings strongly suggest that the true trigger of shame is the prospect or actuality of being devalued by others.

1. Introduction

In 1998, Joseph Dick, a Navy seaman on the USS Saipan, was accused of the rape and murder of Michelle Bosko (Bikel, 2010). He confessed. He was tried and convicted. He served 12 years in prison. Eventually, he expressed how ashamed he was over this act by publicly apologizing to the victim's family.

It seems at first there is no mystery here. No one is surprised when a person found guilty of a crime feels ashamed of what they have done. Such a person has been forced to face his own moral shortcoming and realize his personal failure. Most people would probably feel ashamed in that situation.

The problem? Dick could not have committed the crimes he professed shame over. At the time, he was on duty aboard his ship and could not have left. No physical evidence linked him to the crime. And DNA evidence matched another man, who testified to acting alone. Yet after a great deal of aggressive interrogation by police, who insisted that he was responsible, he gave in and confessed. In the face of so many people insisting he was guilty, he came to feel responsible and ashamed for something he had not actually done.

Joseph Dick was eventually convinced to confess, but even people who maintain their innocence in the face of a wrongful accusation or conviction often feel ashamed. They feel shame merely "because the system has declared [them] publicly guilty" (Wilson, 2002)—even though there is no moral shortcoming or personal failure. Why do people feel shame when others falsely believe they have done something wrong?

1.1. Why do the innocent sometimes feel shame?

One type of explanation for shame comes from *attributional theories of shame*: On this view, shame is activated when two conditions are met: (a) there is an event or outcome that is incongruent with one's representations of one's current or ideal self (e.g., failing an exam, if one aspires to be a good student), and (b) one attributes that event or outcome to one's stable, global self (e.g. blaming that failure on one's low intelligence) (Tangney & Dearing, 2002; Tangney, Stuewig, & Mashek, 2007; Tracy & Robins, 2004). Thus, shame is driven by failure to live up to one's own standards or aspirations, with one's acts reflecting negatively on oneself (Tangney, Stuewig, & Mashek, 2007).

* Correspondence to: T. E. Robertson, College of Business, Stony Brook University, Stony Brook, NY, 11794-3775, United States.
** Correspondence to: D. Sznycer, Department of Psychology, University of Montreal, Montreal, QC, H3C 3J7, Canada.
 E-mail addresses: theresa.robertson@stonybrook.edu (T.E. Robertson), daniel.sznycer@umontreal.ca (D. Sznycer).
[1] These authors share first authorship.

Discussions of shame often contrast it with guilt. In a common view, shame is a public emotion and guilt is a private one. Attributional theories deny this claim (Tangney, Miller, Flicker, & Barlow, 1996; Tangney, Stuewig, & Mashek, 2007; Tracy & Robins, 2006). This denial is based on content analyses of personal accounts of naturally occurring shame and guilt episodes showing that people are no more likely to feel shame rather than guilt when other people are present during the emotional experience, compared to when a person is alone (Tangney, Miller, Flicker, & Barlow, 1996; see also Tracy & Robins, 2006). Instead, according to attributional theories, whereas shame is activated by attributions of negative events to one's *stable, global* self, guilt is activated by attributions of negative events to *unstable, specific* aspects of the self (e.g. blaming the failure on an exam on not having studied enough, rather than not being smart enough) (Tracy & Robins, 2004).

When ashamed, the global self is seen as defective, and this, according to attributional theories, is why the experience of shame is so ugly. The pain of seeing one's self as tainted is so aversive and debilitating that various defensive measures are deployed *in order to avoid the feeling of shame*. These defenses include blaming others, anger, and aggression (Tangney, 1991; Tracy & Robins, 2006), which are reliably correlated with shame (e.g., Fessler, 2001; Tangney, Wagner, Fletcher, & Gramzow, 1992). Other debilitating correlates of shame include anxiety, depression, and paranoid ideations (Gilbert, 2000; Tangney, Wagner, Fletcher, & Gramzow, 1992).

Attributional theories have been used to explain many of shame's observed qualities. However, those theories cannot explain why shame would arise in the absence of wrongdoing—when no personal failure has occurred. That is, they cannot explain our opening puzzle: shame in the innocent.

We suggest that the solution to this puzzle lies in the evolved function of shame. Recent research suggests this emotion is a cognitive system designed by natural selection to limit the likelihood and costs of being devalued due to the spread of negative information about the actor (Sznycer et al., 2016). This threat involves the transient or permanent loss of social attention, socially-derived benefits, or status. People who are not highly valued may fail to receive necessary help or even be active targets of exploitation (Kurzban & Leary, 2001). Ancestrally, being socially devalued would have entailed major fitness costs.

According to the *information threat theory of shame* (Sznycer, 2010; Sznycer et al., 2012, 2016, under review; Sznycer, Schniter, Tooby, & Cosmides, 2015; Sznycer, Cosmides, & Tooby, 2017; Tooby & Cosmides, 2008; see also Gilbert, 1997, 1998; Fessler, 1999, 2007; Schlenker & Leary, 1982), shame will be activated in a person's mind when others learn (or might learn) negative information about that person. By hypothesis, this emotion program is designed to (a) motivate a person to be especially cautious about taking actions that might exacerbate devaluation in an already precarious social situation, (b) limit the spread of potentially damaging information to more people than already know, and (c) limit the costs of any ensuing social devaluation. On this view, the innocent can feel shame if they simply know or suspect that others view them negatively. This is because it is primarily others' beliefs—and not the facts of the matter—that determine a person's reputation and value to others.

1.2. The problem of devaluation

People are selective in whom they associate with and aid, and not all social partners are valued equally (Kurzban & Leary, 2001). In small-scale subsistence societies, as prevailed during human evolution, the potential consequences of losing social benefits are severe. For instance, consider a cross-sectional study of lifetime health problems among the Shiwiar, a hunter-horticulturalist group in the Ecuadorian Amazon. Sixty-five percent of the Shiwiar studied had experienced long-term incapacitating injury or illness at some point in their lives, and had required provisioning from family and friends to survive. If their

families and friends had not valued them enough to provide this aid, all these people would have died (Sugiyama, 2004). However, our zoologically rare abilities and motivations to engage in mutual aid (Ackerman & Kenrick, 2008; Barclay & Willer, 2007) allowed these disabled Shiwiar to recover.

Given the fitness benefits of being valued and the corresponding fitness costs of being devalued, natural selection would have equipped the human mind with a suite of mechanisms for making oneself valuable (Tooby & Cosmides, 1996) and for selectively associating with valuable others. This includes motivations to pursue, acquire, and advertise valued skills (Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013; Sznycer et al., 2017; Tracy, Shariff, & Cheng, 2010), cognitive abilities to seek out especially valuable cooperative partners (Delton & Robertson, 2012; Smith, Pedersen, Forster, McCullough, & Lieberman, 2017), and emotional mechanisms, such as gratitude, for cementing valuable relationships (Algoe & Haidt, 2009; Emmons & McCullough, 2003; Forster, Pedersen, Smith, McCullough, & Lieberman, 2017; Lim, 2010; McCullough, Kimeldorf, & Cohen, 2008; Sznycer, under review). The information threat theory suggests that shame, too, is part of this suite, because it is designed to solve the adaptive problem of being devalued.

Devaluation often happens when a transgression is witnessed or an unwanted personal quality is discovered. However, as we see from the example of the shame experienced by the wrongfully convicted (and the generally poor treatment experienced by the wrongfully convicted, even when their sentences have been overturned; Wilson, 2002), wrongdoing is not necessary for an audience to devalue an individual. For example, noise in the transmission of information or maliciousness may make an audience devalue a person even in the absence of wrongdoing.

1.3. How shame protects against devaluation

How would an evolved mind respond to the threat of devaluation? Being socially devalued entails receiving fewer benefits and incurring more costs from others—a reduction in the prospects of survival and reproduction. Moreover, social devaluation was an ancestrally recurrent situation (e.g., Boehm, 1992; Goodall, 1986; Kurzban & Leary, 2001; Tooby & Cosmides, 1996). Given this adaptive problem, natural selection is expected to have built countermeasures for detecting (and anticipating) social devaluation and for limiting its likelihood and costs. These would include, among others, mechanisms for minimizing the leakage and spread of discrediting information, for improving one's standing on whatever socially valued quality was compromised (e.g. gaining physical strength to make up for the loss of formidability), for bargaining for better treatment, and for acknowledging and tolerating a reduction in status. Besides behavior, the theory also predicts a suite of cognitive, motivational, physiological, and affective responses tailored to the demands of buffering devaluation and coping with the grimmer social landscape resulting from devaluation.

The behaviors shame motivates suggest this emotion is designed to minimize reputational damage. Several different tactics appear to be at work. For instance, shame causes people to avoid eye contact and speaking, and to withdraw from social situations, all of which can prevent damaging common knowledge from forming in others' minds (cf. Thomas, DeScioli, & Pinker, 2018). In the characteristic display of shame, the head is tilted downward and the posture is slumped, which conveys submission and acknowledgement that one's reputation has been impaired (Fessler, 1999; Keltner & Buswell, 1997; Tracy, Robins, & Schriber, 2009). These are potential first steps in being forgiven for perceived transgressions (McCullough, Kurzban, & Tabak, 2013; McCullough, Pedersen, Tabak, & Carter, 2014). Shame also causes people to proactively curry favor with others, by providing costly benefits to those who observed a shameful act (De Hooge, Breugelmans, & Zeelenberg, 2008). Together, these tactics work to lower the probability that others will learn damaging information about a person, and

to reduce the costs a person might experience from those who have acquired damaging information.

1.4. The present research

If the function of shame is to limit the likelihood and the extent of being social devalued (a representational change in the minds of others), then cues predictive of social devaluation will trigger shame. Indeed, if the information threat theory is correct, people should feel shame when they think others hold reputation-damaging beliefs about them—even if the actor knows those beliefs are false. Under the "right" conditions, even an innocent person will feel shame. Attributional theories do not generate this prediction, because innocent people, by definition, have not failed to live up to their own internal standards or aspirations.

We test this hypothesis in two studies. Study 1 uses vignettes to manipulate the likelihood that an audience will devalue the participant. Importantly, in these vignettes participants have done nothing wrong and so any devaluing beliefs in the audience would be *false* beliefs. Study 2 tests face-to-face groups in the laboratory. This study uses social exclusion as a strong experimental manipulation of devaluation. In both studies, shame should track cues of devaluation by others even when the actor has not engaged in any wrongdoing.

2. Study 1

When others believe you have committed a wrong, such as stealing from them, they will value you less as a social partner. According to the information threat theory, these devaluing beliefs of others—even if false—can be sufficient to elicit shame. To test this, we presented a diverse sample of US adults with first-person vignettes. Participants were asked to imagine that they had done nothing wrong. Nonetheless, they were performing an action that could be mistakenly construed as stealing from coworkers. We manipulated how likely the coworkers were to believe that the participant had stolen from them. Does the likelihood of this (false) belief predict shame?

2.1. Method

2.1.1. Participants

Two hundred and seventy-seven US adults (141 women) participated through Amazon Mechanical Turk and were paid $0.45 each. The sample ranged from 18 to 75 years of age ($M = 33$, $SD = 13$). An additional 43 participants (from an original sample of 320) were excluded from analysis for failing to finish the study or for failing an attention check (Chandler, Mueller, & Paolacci, 2014).

2.1.2. Design and vignettes

Participants were asked to imagine that they were performing an action that, though morally unimpeachable, could be mistakenly construed as stealing from coworkers. We manipulated how likely the coworkers were to believe that the participant had stolen from them.

In all conditions, participants were asked to imagine: "You work nights at a bar as a waiter to get an extra income. You work with four other waiters (…) You all put the tips made during the shift in a box, and then you split the money equally among the five of you (…) You realize that you need change," so "you take a $50 bill from your wallet and put it in the tip box." Then the participant reaches into the tip box and pulls out $50 in change (see supplement for full text of vignettes). The vignettes were constructed so that the coworkers witness only the second part of the participant's act of making change (taking smaller-denomination money out of the box), thus potentially being led to believe, mistakenly, that the participant is stealing from the common resource pool. Across three between-subjects conditions, we manipulated how likely coworkers are to (falsely) believe that the participant is stealing from the tip box.

In the *no-belief* condition, the participant makes change alone in the bar; no coworkers are present to form false beliefs. In the *possible-belief* condition, two coworkers walk into the bar just as the participant is taking money out of the box. Because they are searching for a lost cell phone, it is not clear whether they saw the participant taking the money from the box. In this condition, the coworkers could believe the participant is stealing from them, but the participant cannot be sure. In the *likely-belief* condition, when the two coworkers walk in they stare directly at the participant as the latter takes money from the tip box. The information threat theory predicts an increase in the intensity of shame as the likelihood of devaluation increases across these conditions.

2.1.3. Measures

Participants rated their *shame experience* on a single item: "I would feel shame." Participants also rated a set of phenomenological, physiological, and behavioral outputs of shame (*shame behaviors*) using a 9-item scale (based on the existing literature; Fessler, 1999; Tracy, Robins, & Schriber, 2009; Wicker, Payne, & Morgan, 1983). Items included: "I would feel like avoiding eye contact with my coworkers" and "I would feel like leaving the situation immediately" (full measure in supplement). All items were measured on 5-point scales (1 = *not at all*; 5 = *very intensely*). The shame behaviors scale was internally reliable (Cronbach's α = 0.97; confirmatory factor analysis indicated adequate fit for a one-factor model, see supplemental information). Therefore, responses were averaged to form a composite. The single-item shame experience measure and the shame behaviors scale correlated highly ($r = 0.78$, $p < .001$).

2.2. Results

2.2.1. Do participants feel more shame when their coworkers are more likely to (erroneously) believe that they were stealing?

Yes: As the likelihood of devaluing beliefs increased, participants felt more shame ($F(2, 157) = 8.83$, $p = .0002$, $\eta_p^2 = 0.10$; see Fig. 1). This is true, despite the fact that the participant's act, an innocent act, is the same across the three audience belief conditions. Successive means were each significantly or marginally different from each other. The presence of a distracted audience elicited more shame than no audience (possible belief versus no belief: $t(103) = 2.35$, $p = .02$, $r = 0.23$; we use r throughout as a measure of effect size for pairwise comparison, Rosenthal, Rosnow, & Rubin, 2000) and an attentive audience elicited marginally more shame than a distracted audience (likely belief versus possible belief: $t(101) = 1.71$, $p = .09$, $r = 0.17$).

Shame behaviors also increased as the likelihood of devaluing beliefs increased ($F(2, 157) = 21.83$, $p < .0001$, $\eta_p^2 = 0.22$; see Fig. 1). Successive means were again different from each other ($t(103) = 3.69$, $p = .0004$, $r = 0.34$, and $t(101) = 2.69$, $p = .008$, $r = 0.26$).

2.3. Discussion

Participants in this study imagined themselves acting innocently; they were making change, not stealing. Nonetheless, when there was a potential for others to view them as stealing—and hence devalue them—participants reported that they would feel and act ashamed. An attentive audience elicited more shame than a distracted audience, which in turn elicited more shame than the absence of an audience. Importantly, the incremental audience effect was found for an action that appeared to be, but actually was not, theft.

These results are consistent with the information threat hypothesis that shame is modulated by cues of social devaluation. Devaluing beliefs in the minds of others—despite being known to be false—led to feelings of shame and shame-driven responses.

3. Study 2

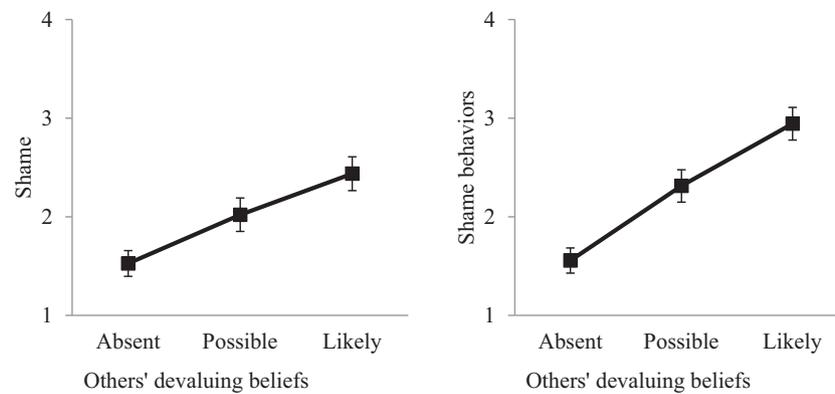Study 1 demonstrated that wrongdoing is not necessary for shame

Fig. 1. Shame and shame behaviors increase as the probability of devaluation increases. Importantly, participants have done nothing wrong; any devaluing beliefs that others might hold would be false beliefs. Error bars represent ± SEM.

activation. This study also showed that shame is modulated in step with cues that predict devaluation, such as the extent to which others are perceived as having access to discrediting information.

However, Study 1 is silent about how much shame is affected by *one's own* disreputable behavior. Perhaps in the absence of a transgression shame is driven by the prospect of devaluation (as in Study 1). But if a person has actually committed a transgression, will shame be driven by the act itself, by others' reactions, or some combination? Recall that according to attributional theories of shame, the proper domain of shame is personal shortcomings and how they reflect upon the self (e.g., Tangney, Stuewig, & Mashek, 2007; Tracy & Robins, 2004). On this view, perhaps the audience effect observed in Study 1 obtained only because we assayed shame beyond its proper domain; perhaps the effect from Study 1 would disappear once a transgression is actually committed. In other words, transgressions, not the beliefs of others, should generally drive shame. On the other hand, the information threat theory of shame argues that information-triggered devaluation is the proper domain of shame and so the effect of publicity will be evident even when there is a transgression (see Smith, Webster, & Eyre, 2002). Study 2 sets these two possibilities against each other and tests the joint effects of the extent of a transgression and cues of devaluation on shame activation.

Study 2 expands on Study 1 in three ways. First, participants interacted in actual face-to-face groups. Second, devaluation was signaled by an experimental manipulation of social exclusion. Participants were ostensibly excluded for failing to contribute to the group in a public good game. In fact, however, exclusion was randomly assigned and was independent of participants' actual contributions. According to the information threat theory, being excluded can elicit shame independently of participants' actual levels of contribution. Finally, we measured perceptions of devaluation to test whether it mediates the relationship between social exclusion and shame.

3.1. Method

3.1.1. Participants

Fifty-seven undergraduate women participated in exchange for $10 (mean age = 19, SD = 1). Seven of them expressed suspicion about the exclusion manipulation; removing them does not change the pattern of results, and they are included in all analyses. In accordance with IRB requirements, participants were prescreened for depression (using the Center for Epidemiological Studies Depression Questionnaire; e.g., Blumenthal et al., 2003; Reisner et al., 2009); f they tested as severely depressed, they were assigned to the inclusion condition and removed from analyses. From an original sample of 68, 11 participants were removed from analyses because prescreening showed they were depressed.

3.1.2. Design

Participants completed a cooperative task with a group they met face-to-face. There were two between-subjects independent variables. The first was a quantitative measure of participants' actual costly contributions to the group in a public good game. The second was an experimental manipulation of exclusion or inclusion by the other group members. Although exclusion was ostensibly due to low contributions from participants, in fact we used false feedback and randomly assigned participants to the inclusion or exclusion condition.

3.1.3. Procedure

Participants worked in groups of 4–6 people. They first completed a face-to-face "get to know you" task to increase the sense that they belonged to a meaningful group (based on Sedikides, Campbell, Reeder, & Elliot, 1999). They were then separated into semi-private cubicles and completed a series of computer-mediated tasks that could earn them tokens. Although the tokens participants earned were partially determined by their actual success at these tasks, they received false feedback that everyone in their group had performed similarly. Thus, participants believed that everyone finished this stage with similar numbers of tokens. They were told each token could be redeemed for $0.02 at the end of the study.

Next, participants completed a public good game, in which they could either keep the tokens they had earned or contribute any number of them to a group account. Tokens contributed were multiplied by a constant and then divided equally among all group members. Multipliers were intended to equalize the per capita return per token across different group sizes. However, due to an experimenter error, they were set to 1.5, 1.6, or 1.7 for groups of 4, 5, or 6 participants, respectively (rather than 1.5, 1.875. and 2.25, respectively). However, there were no significant differences in proportion contributed among the different group sizes (F(2, 49) = 1.80, p = .18, η_p^2 = 0.07, Ms = 0.71, 0.42, and 0.62, respectively, for groups of 4, 5, or 6), and the size of the multiplier was not correlated with proportion contributed (r = −0.17, p = .22). In addition, most people were run in 4 person groups, and none of the analyses are qualitatively affected if other group sizes are removed. Thus, the main analyses include all subjects.

Each token in the group account returns less than one token per person, so it is individually payoff-maximizing to contribute nothing, and this is true regardless of the contribution level of the other group members. However, the total group payoff is maximized when everyone contributes all their tokens. Thus, there is a tradeoff between group interest and individual interest (Dawes, 1980).

After the contribution stage, participants received false feedback that all the other members had contributed 100% of their tokens. Participants contributed on average 67% of their tokens (SD = 37%). (Percentages are reported because the absolute number of tokens varied

somewhat across participants, according to their earnings.) Thus, the typical participant contributed less than they believed their fellow group members had.

Participants then learned that the next task would involve fewer group members. Through the computer they selected the group members they wanted to continue working with and wrote brief explanations for why they did or did not want to continue working with each member. Next, participants received the experimental inclusion/exclusion manipulation. Half of the participants were told that no one wanted to keep working with them; they saw comments that attributed this exclusion to a lack of contribution, such as, "I don't like being with people who won't help." The other half of the participants were told that everyone wanted to keep working with them; they saw positive comments such as, "I liked working with them" (inclusion comments did not make reference to the participant's contribution level). In both conditions, participants did not continue working with their group. In the exclusion condition, this was because of the exclusion. In the inclusion condition, participants were told that because every other group member wanted to keep working with them, for logistical reasons they would have to complete the next part of the experiment alone. Thus, in both conditions participants worked alone; what varied was whether the others excluded and devalued the participant. Our manipulation of exclusion/inclusion succeeded at being independent of participants' own contributions: Whether measured as a percent of their total tokens or as the raw number, the amount of tokens contributed was not correlated with the experimental manipulation ($r(52) = 0.11$, $p = .43$, and $r(52) = 0.03$, $p = .82$, respectively).

Finally, participants rated how ashamed they felt ("I feel ashamed"; 5-point scale; 1 = "Not at All," 3 = "Moderately," 5 = "Very Much"). This item was presented among 14 other distractor emotion items, all presented in random order. Recall that in Study 1, the single-item shame experience measure strongly correlated with the 9-item shame behaviors scale ($r = 0.78$). Participants also reported how devalued they felt using a 7-item scale. Items included: "I feel that the other participants failed to perceive me as a worthy and likeable person" (all items in supplement). All items were measured on 7-point scales (1 = *strongly disagree*; 2 = *disagree*; 3 = *disagree somewhat*; 4 = *neutral*; 5 = *agree somewhat*; 6 = *agree*; 7 = *strongly agree*). The devaluation scale was internally reliable ($\alpha = 0.86$; confirmatory factor analysis indicated adequate fit for a one-factor model, see supplemental information). Therefore, responses were averaged to form a composite. Participants were then debriefed and paid. Because we manipulated the number of tokens earned, at the end of the experiment all participants received the maximum possible payment of $5, in addition to the show-up fee of $5.

3.2. Results

3.2.1. Did participants feel devalued after exclusion?

Yes: Excluded participants felt more devalued than included participants ($t(55) = 4.64$, $p < .0001$, $r = 0.53$; see Fig. 2). This was true even though excluded and included people contributed similarly to the public good, whether measured as percent of tokens earned ($t(50) = -0.79$, $p = .43$, $r = 0.11$) or number of tokens ($t(50) = -1.15$, $p = .25$, $r = 0.16$).

3.2.2. Did participants feel more shame after exclusion?

Yes: Excluded participants felt more shame than included participants ($t(55) = 2.96$, $p = .005$, $r = 0.37$; see Fig. 2). Importantly, the relationship between exclusion and shame was largely unaffected when controlling for participants' actual behavior: In two separate OLS multiple regressions, being excluded still predicted shame when either (a) controlling for participants' total tokens earned and the percentage they contributed ($\beta = 0.35$, $t(51) = 2.54$, $p = .01$, $r = 0.34$) or (b) controlling for total tokens earned and the number of tokens they contributed ($\beta = 0.35$, $t(51) = 2.57$, $p = .01$, $r = 0.34$).
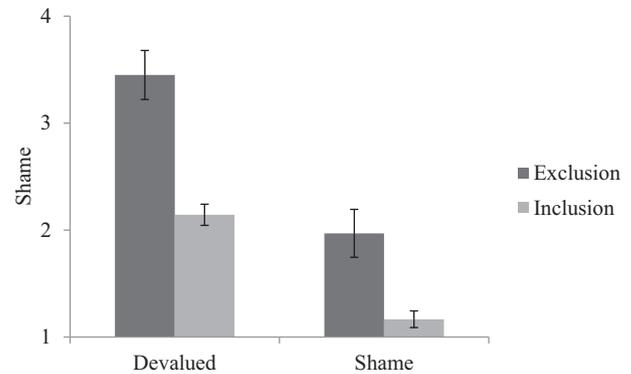

Fig. 2. Excluded participants felt more devalued and ashamed than included participants. Error bars represent ± SEM.

3.2.3. Did perceptions of being devalued statistically mediate between exclusion and shame?

Yes: Based on a bootstrapping mediation analysis (Hayes, 2009), there was an indirect effect of exclusion on shame through perceptions of being devalued (indirect path point estimate = 0.33, given the directional prediction, $p < .05$, one-tailed; see Fig. 3). The direct effect of exclusion on shame was also reduced in size and became non-significant when controlling for perceptions of being devalued (from point estimate = −0.75, $p < .05$, to point estimate = −0.42, ns). This analysis controls for total tokens earned and percentage contributed (the results are similar when controlling for total tokens earned and number contributed). Altogether, these results indicate that participants felt more devalued and ashamed after exclusion, and that participants' perceived devaluation statistically mediates the effect of exclusion on shame.

3.2.4. Did shame vary with contribution level?

Participants' contribution to the public good did not predict shame intensity, whether contribution was measured as percent of tokens contributed ($\beta = -0.03$, $t(51) = -0.18$, $p = .86$, $r = 0.03$) or absolute number of tokens contributed ($\beta = -0.02$, $t(51) = -0.13$, $p = .89$, $r = 0.02$). This was also true within each condition, again for both percent of tokens contributed (Exclusion: $\beta = 0.04$, $t(29) = 0.24$, $p = .81$, $r = 0.04$, Inclusion: $\beta = -0.14$, $t(19) = -0.60$, $p = .56$, $r = 0.14$) or absolute number of tokens contributed (Exclusion: $\beta = 0.08$, $t(29) = 0.42$, $p = .68$, $r = 0.08$; Inclusion: $\beta = -0.08$, $t(19) = -0.35$, $p = .73$, $r = 0.08$; see supplemental information for a table of these results).

We also tested for quadratic effects; such effects may be observed if, for example, shame negatively correlates with contributions among people contributing moderate to high amounts but participants who contribute very little do so because they are relatively immune to shame.[2] However, there was no significant quadratic prediction of shame intensity by contribution level, measured either by percent of
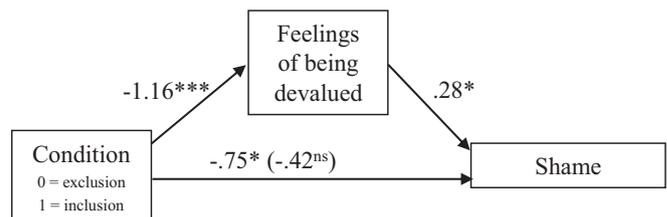

Fig. 3. Participants felt more devalued and ashamed after exclusion. Their perceived level of devaluation statistically mediates the effect of exclusion on shame. Unstandardized regression weights; indirect effect of exclusion on shame through perceptions of being devalued = 0.33. Note: * = $p < .05$, ** = $p < .01$.

tokens contributed or absolute number of tokens contributed (see supplemental information for table of results).

Moreover, when controlling for inclusion/exclusion condition, participants' contributions to the public good did not predict intensity of shame, whether contribution was measured as percent ($\beta = 0.01$, $t(51) = 0.1$, $p = .92$, $r = 0.01$) or number of tokens contributed ($\beta = -0.007$, $t(51) = -0.06$, $p = .96$, $r = 0.01$).

3.2.5. Did even objectively high contributors feel shame after exclusion?

Forty-four percent of the sample contributed all (100%) of their tokens to the public good. Examining this sub-sample provides a particularly strong test of the information threat theory: Contributing everything is clearly not a self-tainting offense. These contributors are not falling short of a moral standard—they could not have contributed more, and others could not have contributed more either, because the stated contribution level of the other group members is also 100%. Still, half of this sub-sample of top contributors was experimentally excluded. A top contributor who is nonetheless apparently excluded for under-contribution should not logically attribute their exclusion to personal wrongdoing. Restricting the analysis to those who contributed 100% of their tokens reduces the statistical power, of course, but we can still ask how the effect size of exclusion (vs. inclusion) on shame among top contributors compares to the corresponding effect size for the sample as a whole (for top contributors, $N = 13$ for excluded participants, $N = 12$ for included ones).

Did exclusion nonetheless lead to more shame among these top contributors? Yes: Even among participants who contributed 100% of their tokens, exclusion led to more shame than inclusion, with a very similar effect size ($r = 0.33$) as for the entire sample ($r = 0.37$). In addition to the effect size replicating, the difference was marginally significant ($t(23) = 1.68$, one-tailed $p = .055$); the marginal significance is not surprising given the small sample size. Excluded top contributors did not engage in wrongful behavior, yet devaluation still elicited shame in them.

3.3. Discussion

Being excluded—a cue of being devalued—led people to feel shame. Moreover, the link between exclusion and shame was mediated by perceptions of being devalued. Exclusion boosted shame controlling for amount contributed, and even among participants who contributed all they had.

It was somewhat surprising to find that contributions were not associated with shame in this sample. One might have expected such an effect (see Smith, Webster, & Eyre, 2002); after all, observers are likely to find low contributors to be less valuable as cooperation partners (Barclay, 2013). However, it could be that those who contributed less thought the amount of tokens they contributed was commensurate with what others contributed.

Nonetheless, the results were consistent with the predictions of the information threat theory of shame. Others' devaluing responses were sufficient to trigger shame, independent of the participants' contribution level, and even when participants contributed everything they had.

4. General discussion

Across two studies, we provided evidence consistent with the information threat theory of shame: Potential or actual devaluation by others is sufficient to elicit shame, independent of the actions one has taken. In Study 1, the innocuous act of making change elicited shame when others were likely to misconstrue it as stealing. In Study 2, exclusion, ostensibly for uncooperativeness in a public good game, elicited shame independently of participants' actual contribution level.

---

[2] We thank an anonymous reviewer for this suggestion.

These findings make sense if the function of shame is to counter the threat of audience devaluation—but not if shame is caused by attributing shortcomings to the self.

These results have implications for existing theories of shame. According to attributional theories of shame, shame is triggered when the individual attributes identity-incongruent events to global, stable aspects of the self (Tangney, Stuewig, & Mashek, 2007; Tracy & Robins, 2004). According to a strong version of this theory, "people focus on others' evaluations because they are feeling shame, not vice versa" (Tangney, Stuewig, & Mashek, 2007, p. 349). The results of the present studies cast doubts on the latter claim: The prospect or actuality of devaluation is sufficient to elicit shame, even when there is no wrongdoing (Study 1) and the act is morally virtuous (Study 2). Reciprocally, the attribution of identity-incongruent events to one's global self does not seem to be a necessary condition for shame to be elicited—it is doubtful that the acts of making change or contributing all of one's endowment to a public good game are incongruent with the identities of most people.

Although we have used the information threat theory of shame as a point of comparison with attribution theories of shame, there are other evolutionarily-oriented theories of shame. Such theories of shame substantially overlap. They agree, for example, that shame is a product of natural selection, that shame is sensitive to other people's evaluations of the self, and that shame motivates remedial behavior (Fessler, 1999, 2007; Gilbert, 1998; Gilbert & McGuire, 1998; Sznycer, 2010; Sznycer et al., 2016). But there are also differences. According to one view (Fessler, 1999, 2007), shame is elicited when the individual violates a norm and "functions to enhance conformity to cultural standards for behavior that form the basis for much cooperation" (Fessler, 2007 p. 174). (Alongside this derived shame system there is a phylogenetically older shame system triggered by the experience of subordinance (Fessler, 1999)). The scope of the information threat theory (Sznycer, 2010; Sznycer et al., 2016), however, is broader than norm-governed cooperation and coordination: Shame should also be triggered by any trait, action, or circumstance that would lead you to be devalued by any individual or set of individuals who can affect your welfare. Further, under the information threat theory, shame functions to limit information-triggered devaluation rather than to enhance conformity.

It is difficult to see how participants who fully cooperated—who contributed 100% of their tokens to the public good—can be construed as having broken a cooperative norm. The fact that excluded participants who fully cooperated felt more shame than included participants who did the same, with a very similar effect size as for the entire sample, sits uneasily with the hypothesis that norm-breaking is necessary to elicit shame. However, the studies reported here were not designed to arbitrate among different evolutionary theories of shame.

We also note that devaluation in the absence of wrongdoing may elicit anger and aggression besides shame (Leary, Twenge, & Quinlivan, 2006; Sell et al., 2017), and that different evolutionary–psychological theories appear to make different predictions regarding the exact variables that may incline the system to issue submissive shame vs. aggressive anger. Thus, while the present results call into question the hypothesis that shame is triggered by attributions of negative events to a defective self, independent of how others view the self, these results do not arbitrate among different evolutionary theories of shame or definitively address the relation between norms and shame. Future work can clarify these issues.

The data presented here strongly suggest that the true trigger of shame is negative perceptions of the self by *others*—not by the self. Shame does lead to perceiving oneself as tainted, as attributional theorists have demonstrated. However, the data suggest that these perceptions have an interpersonal basis (Gilbert, 2000; Leary, Tambor, Terdal, & Downs, 1995; Smith, Webster, & Eyre, 2002; Sznycer, 2010). Shame tracks others' evaluations (Sznycer et al., 2016). Notably, even when a person knows that others' devaluing beliefs are *false*, such false beliefs still trigger shame. This limiting case establishes a lower bound

on the power of others' beliefs: When others' devaluing beliefs are *true*, limiting others' devaluation will be harder and shame will be more intense.

Declarations of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2018.05.010.

References

Ackerman, J. M., & Kenrick, D. T. (2008). The costs of benefits: Help-refusals highlight key trade-offs of social life. *Personality and Social Psychology Review, 12*(2), 118–140.

Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action: The 'other-praising' emotions of elevation, gratitude, and admiration. *The Journal of Positive Psychology, 4*(2), 105–127.

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior, 34*(3), 164–175.

Barclay, P., & Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B: Biological Sciences, 274*(1610), 749–753.

Bikel, O. (2010). The confessions (On Frontline [video transcript]) http://www.pbs.org/wgbh/pages/frontline/the-confessions/etc/transcript.html:PBS.org.

Blumenthal, J. A., Letta, H. S., Babyak, M. A., White, W., Smith, P. K., Mark, D. B., et al. (2003). Depression as a risk factor for mortality after coronary artery bypass surgery. *The Lancet, 362*(9384), 604–609.

Boehm, C. (1992). Segmentary "warfare" and the management of conflict: Comparison of East African chimpanzees and patrilineal–patrilocal humans. In A. H. Harcourt, & F. B. M. deWaal (Eds.). *Coalitions and alliances in humans and other animals* (pp. 137–173). Oxford, England: Oxford University Press.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*, 112–130.

Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology, 104*(1), 103–125.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*, 169–193.

De Hooge, I. E., Breugelmans, S. M., & Zeelenberg, M. (2008). Not so ugly after all: When shame acts as a commitment device. *Journal of Personality and Social Psychology, 95*(4), 933–943.

Delton, A. W., & Robertson, T. E. (2012). The social cognition of social foraging: Partner selection by underlying valuation. *Evolution and Human Behavior, 33*, 715–725.

Emmons, R. A., & McCullough, M. E. (2003). Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology, 84*(2), 377–389.

Fessler, D. M. T. (1999). Toward an understanding of the universality of second order emotions. In A. L. Hinton (Ed.). *Beyond nature or nurture: Biocultural approaches to the emotions* (pp. 75–116). New York: Cambridge University Press.

Fessler, D. M. T. (2001). Emotions and cost/benefit assessment: The role of shame and self-esteem in risk taking. In R. Selten, & G. Gigerenzer (Eds.). *Bounded rationality: The adaptive toolbox* (pp. 191–214). Cambridge, MA: MIT University Press.

Fessler, D. M. T. (2007). From appeasement to conformity: Evolutionary and cultural perspectives on shame, competition, and cooperation. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.). *The self-conscious emotions: Theory and research* (pp. 174–193). New York: Guilford Press.

Forster, D. E., Pedersen, E. J., Smith, A., McCullough, M. E., & Lieberman, D. (2017). Benefit valuation predicts gratitude. *Evolution and Human Behavior, 38*(1), 18–26.

Gilbert, P. (1997). The evolution of social attractiveness and its role in shame, humiliation, guilt and therapy. *British Journal of Medical Psychology, 70*(2), 113–147.

Gilbert, P. (1998). What is shame? Some core issues and controversies. In P. Gilbert, & B. Andrews (Eds.). *Shame: Interpersonal behavior, psychopathology, and culture* (pp. 3–38). New York: Oxford University Press.

Gilbert, P. (2000). The relationship of shame, social anxiety and depression: The role of the evaluation of social rank. *Clinical Psychology & Psychotherapy, 7*, 174–189.

Gilbert, P., & McGuire, M. T. (1998). Shame, status, and social roles. Psychobiology and

evolution. In P. Gilbert, & B. Andrews (Eds.). *Shame: Interpersonal behavior, psychopathology, and culture* (pp. 99–125). New York: Oxford University Press.

Goodall, J. (1986). Social rejection, exclusion, and shunning among the Gombe chimpanzees. *Ethology and Sociobiology, 7*(3–4), 227–236.

Hayes, A. F. (2009). Beyond baron and kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*(4), 408–420.

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122*(3), 250–270.

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin, 127*(2), 187–208.

Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology, 68*(3), 518–530.

Leary, M. R., Twenge, J. M., & Quinlivan, E. (2006). Interpersonal rejection as a determinant of anger and aggression. *Personality and Social Psychology Review, 10*(2), 111–132.

Lim, J. (2010). *Welfare tradeoff ratios and emotions: Psychological foundations of human reciprocity.* Santa Barbara: University of California (Unpublished doctoral dissertation).

McCullough, M. E., Kimeldorf, M. B., & Cohen, A. D. (2008). An adaptation for altruism? The social causes, social effects, and social evolution of gratitude. *Current Directions in Psychological Science, 17*, 281–285.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences, 36*(1), 1–58.

McCullough, M. E., Pedersen, E. J., Tabak, B. A., & Carter, E. C. (2014). Conciliatory gestures promote forgiveness and reduce anger in humans. *Proceedings of the National Academy of Sciences, 111*(30), 11211–11216.

Reisner, S. L., Mimiaga, M. J., Skeer, M., Bright, D., Cranston, K., Isenberg, D., et al. (2009). Clinically significant depressive symptoms as a risk factor for HIV infection among black MSM in Massachusetts. *AIDS and Behavior, 13*(4), 798–810.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach.* Cambridge University Press.

Schlenker, B. R., & Leary, M. R. (1982). Social anxiety and self-presentation: A conceptualization model. *Psychological Bulletin, 92*, 641.

Sedikides, C., Campbell, W. K., Reeder, G. D., & Elliot, A. J. (1999). The relationship closeness induction task. *Representative Research in Social Psychology, 23*, 1–4.

Sell, A., Sznycer, D., Al-Shawaf, L., Lim, J., Krauss, A., Feldman, A., ... Tooby, J. (2017). The grammar of anger: Mapping the computational architecture of a recalibrational emotion. *Cognition, 168*, 110–128.

Smith, A., Pedersen, E. J., Forster, D. E., McCullough, M. E., & Lieberman, D. (2017). Cooperation: The roles of interpersonal value and gratitude. *Evolution and Human Behavior, 38*(6), 695–703.

Smith, R. H., Webster, J. M., & Eyre, H. L. (2002). The role of public exposure in moral and nonmoral shame and guilt. *Journal of Personality and Social Psychology, 83*(1), 138–159.

Sugiyama, L. S. (2004). Illness, injury, and disability among Shiwiar forager-horticulturalists: Implications of health-risk buffering for the evolution of human life history. *American Journal of Physical Anthropology, 123*, 371–389.

Sznycer, D. (2010). *Cognitive adaptations for calibrating welfare tradeoff motivations, with special reference to the emotion of shame.* Santa Barbara: University of California (ProQuest Dissertations and Theses Unpublished doctoral dissertation).

Sznycer, D. The emotion–valuation constellation: Multiple emotions are governed by a common grammar of social valuation (Under review).

Sznycer, D., Al-Shawaf, L., Bereby-Meyer, Y., Curry, O. S., Smet, D. D., Ermer, E., et al. (2017). Cross-cultural regularities in the cognitive architecture of pride. *Proceedings of the National Academy of Sciences, 114*(8), 1874–1879.

Sznycer, D., Cosmides, L., & Tooby, J. (2017). Adaptationism carves emotions at their functional joints. *Psychological Inquiry, 28*(1), 56–62.

Sznycer, D., Schniter, E., Tooby, J., & Cosmides, L. (2015). Regulatory adaptations for delivering information: The case of confession. *Evolution and Human Behavior, 36*(1), 44–51.

Sznycer, D., Takemura, K., Delton, A. W., Sato, K., Robertson, T., Cosmides, L., & Tooby, J. (2012). Cross-cultural differences and similarities in proneness to shame: An adaptationist and ecological approach. *Evolutionary Psychology, 10*, 352–370.

Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S., & Halperin, E. (2016). Shame closely tracks the threat of devaluation by others, even across cultures. *Proceedings of the National Academy of Sciences, 113*(10), 2625–2630.

Sznycer, D., Xygalatas, D., Agey, E., Alami, S., An, X.-F., Ananyeva, K. I., Atkinson, Q. D., Broitman, B. R., Conte, T. J., Flores, C., Fukushima, S., Hitokoto, H., Kharitonov, A. N., Onyishi, C. N., Onyishi, I. E., Romero, P. P., Schrock, J. M., Snodgrass J. J., Sugiyama, L. S., Takemura, K., Townsend, C., Zhuang, J.-Y., Aktipis, C. A., Cronk, L., Cosmides, L., & Tooby, J. Cross-cultural invariances in the architecture of shame. (Under review).

Tangney, J. P. (1991). Moral affect: The good, the bad, and the ugly. *Journal of Personality and Social Psychology, 61*(4), 598–607.

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt.* New York: The Guilford Press.

Tangney, J. P., Miller, R., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology, 70,* 1256–1269.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology, 58*, 345–372.

Tangney, J. P., Wagner, P. E., Fletcher, C., & Gramzow, R. (1992). Shamed into anger? The relation of shame and guilt to anger and self-reported aggression. *Journal of Personality and Social Psychology, 62*(4), 669–675.

Thomas, K. A., DeScioli, P., & Pinker, S. (2018). Common knowledge, coordination, and the logic of self-conscious emotions. *Evolution and Human Behavior, 39*(2), 179–190.

Tooby, J., & Cosmides, L. (1996). Friendship and the Banker's Paradox: Other pathways to the evolution of adaptations for altruism. In W. G. Runciman, J. Maynard Smith, & R. I. M. Dunbar (Vol. Eds.), *Evolution of social behaviour patterns in Primates and man. Proceedings of the British Academy*. *Vol. 88. Evolution of social behaviour patterns in Primates and man. Proceedings of the British Academy* (pp. 119–143).

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.). *Handbook of emotions* (pp. 114–137). (3rd Ed.). (NY: Guilford).

Tracy, J. L., & Robins, R. W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry, 15*(2), 103–125.

Tracy, J. L., & Robins, R. W. (2006). Appraisal antecedents of shame and guilt: Support for a theoretical model. *Personality and Social Psychology Bulletin, 32*(10), 1339–1351.

Tracy, J. L., Robins, R. W., & Schriber, R. A. (2009). Development of a FACS-verified set of basic and self-conscious emotion expressions. *Emotion, 9*(4), 554–559.

Tracy, J. L., Shariff, A. F., & Cheng, J. T. (2010). A naturalist's view of pride. *Emotion Review, 2*(2), 163–177.

Wicker, F. W., Payne, G. C., & Morgan, R. D. (1983). Participant descriptions of guilt and shame. *Motivation and Emotion, 7*(1), 25–39.

Wilson, J. (2002). A perpetual battle of mind [Electronic Version] (Frontline (online material), from) http://www.pbs.org/wgbh/pages/frontline/shows/burden/cameras/memo.html.