Original Article

Common knowledge, coordination, and the logic of self-conscious emotions

Kyle A. Thomas [a,*], Peter DeScioli [b], Steven Pinker [a]

[a] *Department of Psychology, Harvard University, United States*
[b] *Department of Political Science, Stony Brook University, United States*

ARTICLE INFO

1. Introduction

Imagine spilling a plate of food into your lap in front of a crowd. Afterwards, you might fix your gaze on your cell phone to avoid acknowledging the bumble to onlookers. Similarly, after disappointing your family or colleagues, it can be hard to look them in the eye. Why do people avoid acknowledging faux pas or transgressions that they know an audience already knows about?

Following a transgression, people feel the negative self-conscious emotions of shame, embarrassment, or guilt, and these emotions help them regulate their relationships (Beer, Heerey, Keltner, Scabini, & Knight, 2003; Keltner & Buswell, 1997; Ketelaar & Au, 2003; Miller, 1995; Steckler & Tracy, 2014; Sznycer et al., 2016; Tangney & Tracy, 2012; Tracy & Robins, 2004). A transgressor has displayed ineptitude, which can damage his reputation as a valuable cooperator, or a disregard for someone's welfare, which can damage his reputation as a trustworthy cooperator. The discomfort caused by the resulting emotions, even when privately felt, motivates a person to manage these threats by drawing his attention to the transgression and motivating him to make amends and avoid similar acts in the future (Ketelaar & Au, 2003; McCullough, 2008; Sznycer et al., 2016; Trivers, 1971).[1]

The idea that self-conscious emotions regulate relationships also explains why the presence of an audience intensifies feelings of embarrassment, shame, and guilt (Bosch et al., 2009; Gruenewald, Kemeny,

Aziz, & Fahey, 2004; McCullough, 2008; Modigliani, 1971; Smith, Webster, Parrot, & Eyre, 2002; Tangney, Miller, Flicker, & Barlow, 1996; Wolf, Cohen, Panter, & Insko, 2010). If onlookers infer that a transgression is the result of a stable disposition that predicts future incompetence or exploitation, they now have reason to devalue, ostracize, or punish the transgressor. To prevent these damaging consequences, the transgressor must persuade the onlookers either that the act was not intentional and hence unrepresentative of his underlying disposition, or that he will change his disposition and will not repeat the behavior in the future. Moreover, for such assurances to be more than self-serving cheap talk, they must be made credible: The transgressor must endure a cost, in the form of visible discomfort and perhaps tangible restitution, and display signs that the changed priorities are products of involuntary emotions rather than conscious strategic calculations. Indeed, research on the psychology of contrition and forgiveness shows that the negative self-conscious emotions have these specifications (Dijk, de Jong, & Peters, 2009; Dijk, Koenig, Ketelaar, & de Jong, 2011; Frank, 1988; Keltner & Buswell, 1996; Ketelaar & Au, 2003; McCullough, 2008; McCullough, Kurzban, & Tabak, 2010; Tracy & Robins, 2004; Trivers, 1971).

As we noted at the outset, intuition suggests that self-conscious emotions have an additional feature: They are felt more intensely still when the actor acknowledges to the audience that he is aware of their knowledge of his transgression (say, by meeting their gaze). In this paper we attempt to make this hypothesis precise, test whether it is correct, and attempt to explain it.

1.1. The game theory and psychology of common knowledge

The difference between evading and acknowledging unflattering events parallels the distinction in game theory between *shared* and *common knowledge*. Common knowledge is an infinite recursion of

---

* Corresponding author at: Department of Psychology, Harvard University, William James Hall 964, 33 Kirkland Street, Cambridge, MA 02138, United States.
E-mail address: kylethomas@post.harvard.edu (K.A. Thomas).

[1] An analysis of knowledge of social transgressions requires precisely differentiating the individual parties in dyadic interactions. The use of *he or she* or generic *they* in such discussions can quickly become unwieldy or ambiguous (Pinker, 2014). To avoid confusion, we will consistently refer to a hypothetical actor using a masculine singular pronoun, and to hypothetical onlookers using feminine or plural pronouns (Pinker, 2014).

social knowledge states, such that A knows x, B knows x, A knows that B knows x, B knows that A knows x, A knows that B knows that A knows x, A knows that B knows that A knows that B knows x, ad infinitum. In contrast, shared knowledge refers to any lesser, finite level of social knowledge states (e.g., A knows that B knows that A knows x, but nothing more).

Common knowledge plays an important role in *coordination games*—interactions with multiple equilibria in which each player's best move depends on the other player's move (Lewis, 1969; Schelling, 1960; Skyrms, 2004). The paradigm case is the Stag Hunt, in which two players can individually hunt rabbits (a low payoff) or can work together to hunt stag (a high payoff), but if one player hunts stag alone then the hunt fails and that player earns nothing. To coordinate, each player needs to know that the other player knows there is an opportunity to hunt stag, and further, that this information is common knowledge. Lesser forms of shared knowledge, in which each player knows that stag are running but wonders whether the other player knows that, or knows that he knows it, can easily lead to costly discoordination (Chwe, 2001; Dalkiran, Hoffman, Paturi, Ricketts, & Vattani, 2012; Rubinstein, 1989; Schelling, 1960).

Recent psychological research shows that people do in fact represent common knowledge as qualitatively distinct from shared knowledge and their strategic decisions differ based on this distinction (Thomas, DeScioli, Haque, & Pinker, 2014). Thomas et al. (2014) placed participants in hypothetical scenarios with the payoffs of a Stag Hunt: They had to decide whether to work alone for a certain but lower payoff, or to try to work together with a partner for a higher payoff that they only received if the partner made the same choice. The source of the information about the payoffs varied, sometimes allowing only private knowledge, sometimes shared knowledge (conveyed by an unreliable messenger boy), sometimes common knowledge (conveyed by a loudspeaker). Participants were most likely to choose to coordinate, and thus to attain the higher payoff, when they had common knowledge of the opportunity, in line with game-theoretic predictions. Of course, no one can explicitly represent an infinitely embedded proposition; presumably common knowledge is represented in people's minds as a finite mental formula or symbol, corresponding to the intuition that something is "public" or "out there".

People's concept of common knowledge affects not only their strategic reasoning but also a variety of other psychological phenomena including indirect speech (Chakroff, Thomas, Haque, & Young, 2015; Lee & Pinker, 2010), moral judgment (DeScioli & Kurzban, 2013), pluralistic ignorance (Willer, Kuwabara, & Macy, 2009), and diffusion of responsibility (Thomas, De Freitas, DeScioli, & Pinker, 2016). We investigate the role of common knowledge in the self-conscious emotions. Specifically, we propose that compared to shared knowledge, common knowledge of a transgression imposes a greater imperative to rectify it, which in turn causes more intensely felt self-conscious emotions.

1.2. Social transgressions, social signals, and shared knowledge

We start with the idea that an offense poses a greater threat to the transgressor's reputation if an audience not only privately knows about the transgression but also knows that the transgressor knows they know. Someone who botches a job or lets down a friend will look bad if he does not apologize, but will look even worse if the audience knows he knows the audience viewed the offense, and still no apology is forthcoming.

The reasoning is as follows. A transgressor who fails to apologize when an audience observes the offense, without shared knowledge, is, to be sure, already in a compromised position. But the audience could still be uncertain about the transgressor's intentions, so from their point of view, it's still possible that the transgressor intends to behave differently in the future. This uncertainty gives observers reason not to respond draconianly (since they too may lose the benefits of future cooperation with the actor), and such a moderate pressure on the

transgressor should give rise to mild forms of embarrassment, guilt, and shame.

In contrast, a transgressor who fails to apologize even when observers know that he knows that they know of the offense has transmitted additional information: that he is unwilling or unable to change his behavior to preserve his standing with the observers. In other words, he would be signaling that he is prepared to disengage from cooperation entirely and forgo all of its benefits: the strategy of a psychopath, lone wolf, oddball, or loose cannon (see also Goffman, 1959, 1978). For a member of a highly social species, the consequences of such a display can be severe, and so we suggest that the self-conscious emotions are adapted to prevent such consequences.

Furthermore, research has shown that transgressors tend to apologize and confess strategically (Sznycer, Schniter, Tooby, & Cosmides, 2015). Specifically, transgressors tend to confess and apologize when doing so most benefits them, that is when: (1) the cost of the transgression to the victim is low; (2) the benefit of the transgression to them is high; and (3) victims may already know of, or may easily discover the transgression otherwise. Victims thus face a signal-detection problem when they receive an apology—Is it an honest signal or strategic cheap talk? Importantly, when the victim thinks that the transgressor thinks he is revealing the transgression to the victim for the first time, instead of the victim already knowing what the transgressor did, then the transgressor's apology appears more credible, since it reveals the offense rather than only patching things up after the offense was already uncovered (Sznycer et al., 2015; Utikal, 2012; Weiner, Graham, Peter, & Zmuidinas, 1991). Thus, as soon as a transgression becomes shared knowledge, both parties could reasonably surmise that the victim's expectation of a forthcoming apology increases, and so the transgressor's time window for providing one narrows accordingly. In such cases, more than only stating an apology may be required to make it credible, and the uncontrollable emotional signals produced by self-conscious emotions may increase the credibility of what otherwise might be perceived as mere cheap talk.[2]

This reputation-management theory of negative self-conscious emotions is consistent with the fact that these emotions are evoked by two distinct types of transgression: those that show disregard for someone's welfare and those that reveal incompetence. Both categories follow from theories of cooperation and partner choice, which show how individuals gain more from cooperation when they prefer partners who both value their welfare and are competent enough to reciprocate (Tooby & Cosmides, 1996; Trivers, 1971). Transgressors who fail to make amends are more likely to re-offend in the future, whether because of indifference, incompetence, hostility, or ignorance. We suggest that a lack of contrition is even more diagnostic when the audience knows the actor knows that the audience viewed the transgression.

1.3. Relationships, coordination, and common knowledge

Further, we propose that common knowledge of a transgression is even worse for the actor than shared knowledge. Shared knowledge of an offense threatens to damage an actor's reputation; common knowledge adds the additional threat of destabilizing a valuable relationship.

From a game-theoretic perspective, relationships can be viewed as coordination games (Dalkiran et al., 2012; Lee & Pinker, 2010; Pinker, Nowak, & Lee, 2008), as it is often mutually advantageous to show kindness, respect, support, indifference, affection, or malice toward those who show the same to us. That is, social partners generally benefit by coordinating on the same type of relationship, and relationship mismatches can be costly. People can use present signals and past histories to establish cooperative relationships. Even so, either party can unilaterally abandon a relationship at any time. Thus, each partner benefits by

---

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

periodically checking the status of the relationship and reaffirming his understanding of the relationship to his counterpart.

Because they rely on synchronized beliefs and expectations, relationships are fragile and can be undermined by intentional or even accidental transgressions. A friendship can turn cold, and strangers may transition from neutral to antagonistic. In these cases, individuals are in the difficult situation of guessing each other's mental states to predict their future behavior, and an unacknowledged transgression can disrupt a relationship by tipping a partner's expectations toward noncooperation in the future.

This problem is well studied in the context of violent conflict where rivals can inadvertently enter a "Hobbesian trap", such as an escalating arms race (Pinker, 2011; Schelling, 1960). For example, two nations without close ties might prefer to avoid war after a dispute, but they each mobilize military forces in case the other decides to attack, which in turn incentivizes each to attack preemptively to avoid losing the advantage of striking first.

This same dynamic can occur in interpersonal relationships. A friend, spouse, colleague, or stranger might turn a cold shoulder simply because he fears the same from his counterpart. When a public transgression occurs—stolen money, suspected infidelity, a misspoken slur, or a spilled plate of food—individuals might second-guess their relationships and end up in a kind of "defection trap" in which they each view the relationship more negatively solely because they think their counterpart has done the same.

To escape these traps, partners need some way to reaffirm the relationship by reassuring each other that they both intend to maintain it. We propose that the self-conscious emotions are designed to facilitate these reaffirmations: They help a transgressor to immediately and publicly signal contrition in order to prevent a spiraling deterioration of the relationship. Consistent with this idea, research has shown that embarrassment and blushing can serve as signals of contrition and help appease a potentially hostile audience (Dijk et al., 2009; Dijk et al., 2011; Keltner, 1995; Keltner & Buswell, 1997), and all three negative self-conscious emotions motivate people to take corrective action and make amends (Beer et al., 2003; Keltner & Buswell, 1997; Ketelaar & Au, 2003; Tracy & Robins, 2004).

Importantly, the logic of coordination implies that common knowledge of a transgression can be more damaging to a relationship than even arbitrarily high levels of shared knowledge. By aligning individuals' beliefs and expectations, common knowledge of a transgression can force social partners to acknowledge uncomfortable truths, and create a defection trap that may destroy their relationship if left unaddressed. In contrast, if a social transgression is merely shared knowledge, social partners are not necessarily forced to acknowledge it, which can keep them out of a defection trap. A small amount of uncertainty means that a failure to apologize is not necessarily a direct affront challenging the status quo, but might instead be an attempt to evade the offense. Although the victim might still prefer apology over evasion, the latter at least suggests that the actor intends to maintain the status quo. In short, common knowledge of a transgression forces people to revisit, renegotiate, and/or reaffirm the terms of their relationship; whereas, even high levels of shared knowledge provide enough wiggle room to allow them to simply overlook the violation if they so desire (Dalkiran et al., 2012 provide a formal model of this phenomenon).

Finally, the threats posed by common knowledge may also go beyond the relationship itself to include third-parties who observe or find out about the violation. When a violation is common knowledge, third parties are better able to coordinate condemnation and punishment of the violator (DeScioli & Kurzban, 2013). Hence, common knowledge poses a special threat from third parties in addition to the damage it can do within relationships. Here we focus on dyadic relationships but we will return to the issue of third-party coordination in the General Discussion (Section 4.3).

In sum: a cooperative relationship is threatened by a transgression that is common knowledge (unless it is specifically redressed), but less so by a transgression that is merely shared knowledge (which need not necessarily be redressed).

1.4. How do self-conscious emotions help individuals manage relationships?

Predominant theories of the self-conscious emotions maintain that self-conscious emotions function to help individuals manage relationships and navigate their social environment (see Beer et al., 2003; Dijk et al., 2009; Keltner, 1995; Keltner & Buswell, 1997; Ketelaar & Au, 2003; McCullough, 2008; Miller, 1995; Steckler & Tracy, 2014; Sznycer et al., 2016; Tangney & Tracy, 2012; Tracy & Robins, 2004). Specifically, these theories focus on how the self-conscious emotions motivate people to avoid transgressions, prevent harmful information from spreading, and appease others after a transgression occurs. The finding that the self-conscious emotions are felt more intensely when an audience is present provides empirical support for this set of ideas.

However, while maintaining one's reputation is one important aspect of managing relationships, our analysis suggests another critical element, which existing theories have not addressed: establishing and maintaining coordination. If relationships require coordination, then the self-conscious emotions will be felt more intensely when a transgression is acknowledged to an audience, compared to when it is left unacknowledged, because common knowledge of the offense poses a threat to coordination.

Hence, the coordination theory of the self-conscious emotions predicts that common knowledge of a transgression will elicit more intense feelings of embarrassment, shame, or guilt than if the transgression is merely shared knowledge between an audience and transgressor. The reputation theories proposed in previous research do not clearly predict this additional effect of common knowledge. A transgressor's reputation can be harmed as soon as someone else learns about his transgression. Thus, knowing that an audience is aware of one's transgression is sufficient to know one's reputation is in peril. Common knowledge need not increase this peril because once a transgressor knows a third party is aware of the offense and yet still fails to address it, the reputational damage has already been done. We propose that because the self-conscious emotions track concerns about both reputation and coordination, they should be felt more intensely with shared knowledge than private knowledge, and more intensely still with common knowledge than with shared knowledge.

1.5. Empirical tests

We test the coordination theory in two experiments, one using hypothetical scenarios and another using a real, embarrassing situation in which participants sing to an audience of strangers. A problem in testing any theory of the negative self-conscious emotions of embarrassment, guilt, and shame is how to distinguish them from other emotions with a negative valence such as anger, sadness, or anxiety. One of the oldest findings in affective science is that the different emotions are arrayed along a small number of dimensions, of which positive-negative is one of the most salient (Rubin & Talarico, 2009; Russell & Barrett, 1999; Watson, Clark, & Tellegen, 1988). Furthermore, the different emotions at the negative pole fall into a natural class, and may share underlying phenomenology, physiological substrates, and pathologies (as in the comorbidity of depression, anxiety, and social phobias) (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001). Thus, in testing a hypothesis that seeks to identify the cognitive triggers and behavioral sequelae of specific emotions, it is essential to distinguish the experience and reporting of those emotions from a more global negativity that may be elicited by a variety of nonspecific stressful situations. In both experiments, we thus attempt to distinguish self-conscious emotions from more general negativity by subtracting a composite of participants' reported negative emotions from their reported self-conscious emotions.

In the first experiment, participants read first-person vignettes describing situations with a transgression that typically elicits either negative self-conscious emotions, anger, or sadness. We chose anger and sadness as cases of well-studied basic emotions (Ekman, 1999; Tracy & Robins, 2004). In different versions of the scenarios, we varied the characters' levels of knowledge about the transgression as private, shared, or common knowledge. Participants were asked to imagine themselves in the scenario and then rate how much they would feel different emotions.

Importantly, we note that participants' ratings of emotions for hypothetical scenarios could reflect a mixture of the actual emotions participants may be experiencing vicariously, as well as their beliefs about which emotions they anticipate they would feel in such scenarios (which may or may not be accurate). Written scenarios can potentially evoke genuine emotions. For instance, people commonly report cringing, fear, and other emotional reactions when empathizing with a fictitious character in a stressful situation. Such vicarious and imagined emotions are the basis for imagery-based therapies such as desensitization and implosion (Holmes, Arntz, & Smucker, 2007; Mar & Oatley, 2008). However, we cannot distinguish genuine from anticipated emotions from self-reports, although we expect them to be highly correlated.

If, as predicted by the coordination theory, acknowledging a transgression is a particularly potent trigger for self-conscious emotions, then participants will report more intense self-conscious emotions when transgressions are common knowledge compared to shared knowledge. As explained above, the coordination theory goes further than predicting that self-conscious emotions should be sensitive to these varying knowledge levels; it also predicts that these effects will be relatively specific to the self-conscious emotions, as opposed to other negative emotions that serve additional important functions besides any role they might play in managing relationships. The coordination theory thus predicts that knowledge levels: (a) will affect self-conscious emotion ratings after subtracting out general negative affect; and, (b) will affect self-conscious emotions more than the traditional basic emotions of anger and sadness.

The second experiment then goes further by actually placing participants in an embarrassing situation—karaoke. Participants sang in a karaoke performance for a panel of judges in a separate room connected by a live two-way video feed, and were either told that the judges were aware that the participant knew the judges were watching through the camera (common knowledge), or were told that the judges thought they were secretly observing the performance, unbeknownst to the participant (shared knowledge). Participants then reported how intensely they felt 21 different emotions during their performance. The coordination theory predicts greater embarrassment with common knowledge than shared knowledge, and more specifically, participants' distinctive feelings of embarrassment, after subtracting general negative affect, will be greater with common knowledge than shared knowledge.

2. Experiment 1

In this experiment participants reported how they would react to different levels of knowledge in one of five scenarios. Three of the scenarios were designed to evoke negative self-conscious emotions: getting caught making fun of a friend, audibly passing gas during a lecture, or dishonestly reporting purchases for reimbursement. The other two were designed to evoke non-self-conscious negative emotions: sadness at discovering that one's romantic partner was about to end the relationship, and anger at discovering that an acquaintance was trying to sabotage a valued friendship. For each scenario, participants rated how much they would feel a battery of emotions. Ratings of multiple emotions allowed us to examine self-conscious emotions after subtracting general negative affect, distinguishing the former from the latter.

Each participant read five different vignettes (presented in randomized order) that described different levels of knowledge of the focal event in one of the scenarios. That is, the scenario was varied between subjects; whereas, knowledge was varied within subjects, ranging from *private knowledge*, in which the participant alone had knowledge of the focal event, to *shared knowledge*, in which the participant and another person knew about the focal event but neither was fully apprised of the other's knowledge, to *common knowledge*, in which the focal event was commonly known between the participant and another person.

Because negative self-conscious emotions are typically caused by one's own actions, whereas other negative emotions tend to be evoked by other people's actions, the focal events in the corresponding scenarios had to differ: Participants assumed the role of the actor in the self-conscious scenarios and of an observer in the basic emotion scenarios. As Tangney and Tracy (2012) explain this distinction, self-conscious emotions differ from other kinds of emotions because "...[the self-conscious emotions] fundamentally involve people's reactions to their own characteristics or behavior" (p. 446). In contrast, it is hard to think of doing something oneself that *directly* evokes non-self-conscious emotions such as anger or sadness (as opposed to initiating a chain of unpredictable events, the last of which may be the proximate cause of the emotion).

2.1. Method

2.1.1. Participants

According to a power analysis for detecting medium effects, one hundred participants from the United States were recruited from Amazon's Mechanical Turk for each of the five scenarios ($N = 500$, aged 18–81), and paid $2–$3 for their participation. Because attentiveness is a particular concern with remote participants, comprehension and attention checks were included, and any participants who failed any of these items were excluded (see Procedure), yielding a final sample of 361 participants (55% female, $M_{age} = 34.1$, $SD_{age} = 11.1$).[3]

2.1.2. Procedure

Participants read instructions explaining the task and were given definitions of the six emotions, derived from standard dictionaries, that they would be asked to rate in the study:

- *Anger*—A strong feeling of displeasure, grievance, annoyance, and hostility.
- *Embarrassment*—A feeling of acute and painful self-consciousness, uncomfortableness, awkwardness, and humiliation.
- *Fear*—An unpleasant feeling of anxiety or apprehension caused by the presence or anticipation of danger.
- *Guilt*—An awareness of having done wrong or failed in an obligation, accompanied by feelings of remorse and regret.
- *Sadness*—A feeling of unhappiness, grief, and sorrow.
- *Shame*—A painful negative emotion that combines feelings of dishonor, unworthiness, distress, and humiliation, and is caused by the awareness of having done something wrong or foolish.

To ensure they understood the emotion words, participants were then given the six definitions and asked to select the corresponding emotion. Participants who answered any of these questions incorrectly were excluded from the analyses.

Participants then read five versions of one scenario, each with a different level of knowledge about the focal event, presented in a random

---

[3] We note that in both experiments we report all measures, manipulations, and exclusions in the experiments. Furthermore, all exclusion criteria and statistical analyses were planned before the analysis was conducted, and no data collection was continued after the analyses began.

order. After each version, participants used slider bars to report how strongly they would experience each of the six emotions on a scale of 0–100, and how likely and intensely they would exhibit 16 physical reactions, also on a scale of 0–100. These reactions were drawn from previous emotion research and included: avoid gaze, blanch, blush, clench teeth, cover or touch face, cry, frown, glare, hang head down, nervous laughter, open eyes wide, raise eyebrows, sheepish grin, slump shoulders, stammer, and tremble (see Ekman, 2003; Keltner & Buswell, 1997; Tracy & Robins, 2004).

Participants then reported basic demographic information and responded to a simple attention check, disguised as a final question which asked them to report which emotions from a list of 20 they were experiencing. There was an instruction buried in the prompt that asked them to select "only the enthusiastic option" to show that they were paying attention; participants who did not select it were excluded from the analyses. They were then debriefed and paid.

The following three scenarios were designed to elicit self-conscious emotions:

○ *Teasing*—Participants imagined talking to a friend and mocking a mutual friend's speech impediment, when the person being mocked unexpectedly shows up.
○ *Flatulating*—Participants imagined audibly flatulating in a crowded lecture hall.
○ *Falsifying*—Participants imagined submitting a dishonest reimbursement request to a club, which is discovered by their housemate.

The following two scenarios were designed to elicit either anger or sadness:

○ *Breakup*—Participants imagined discovering that their romantic partner was going to break up with them by overhearing a voicemail left by the partner's friend.
○ *Sabotage*—Participants imagined overhearing an acquaintance trying to sabotage their relationship with a good friend.

Each participant read five different versions of a given scenario, each with a different level of knowledge about the focal event. Because participants assumed the role of the actor in the self-conscious scenarios but the observer in the basic emotion scenarios, the knowledge levels in the two kinds of scenarios were not identical. The self-conscious emotion scenarios were presented with each of the following knowledge levels:

1. *Private knowledge 1*—Only the participant knows about his transgression.
2. *Private knowledge 2*—An observer is aware of the transgression, but the participant doesn't realize this.
3. *Shared knowledge 1*—The participant knows that an observer knows about the transgression, but the observer is unaware he knows this.
4. *Shared knowledge 2*—The participant knows that an observer knows about the transgression, the observer knows the participant knows this, but the observer is unaware that the participant knows the observer knows this.
5. *Common knowledge*—The transgression is commonly known (i.e., the participant knows the observer knows about it, knows the observer knows he knows, ad infinitum).

The basic emotion scenarios were presented with each of the following knowledge levels:

1. *No knowledge*—The participant is unaware of the focal event.
2. *Private knowledge 1*—The participant is aware of the focal event, but the other person doesn't realize this.

3. *Private knowledge 2*—The participant is aware of the focal event, and the other person knows this, but the participant doesn't know the other person is aware that he knows.
4. *Shared knowledge*—The participant is aware of the focal event, and the participant knows that the other person is aware that he knows about it, but the other person is unaware he knows this.
5. *Common knowledge*—The focal event is commonly known (i.e., the other person knows the participant knows about it, knows the participant knows she knows, ad infinitum).

The knowledge levels were not described in these schematic ways (which would be difficult for participants to keep track of) but in concrete descriptions of what the actors looked at, noticed, or failed to notice. For example, in the Shared Knowledge 2 condition of the Teasing scenario, in which the participant imagines making fun of Lisa's speech impediment to Robert, the description reads, "Robert laughs at the inside joke just as Lisa appears out of a nearby store. She hears you make fun of her impediment and her facial expression changes. You try not to make eye contact with her by quickly looking in the opposite direction; however, you realize she managed to catch you in the act of looking away. Yet, by this time, you had already turned far enough that it does not appear to her that you noticed that she saw you." The full text of all vignettes is supplied in Supplementary Materials 2.

2.1.3. Data analysis

Although shame, embarrassment, and guilt are distinct emotions, they overlap enough that people often conflate them when reporting their feelings (Keltner & Buswell, 1996; Tangney et al., 1996; Tangney & Tracy, 2012). In a preliminary analysis, we confirmed that participants' ratings of embarrassment, shame, and guilt were highly correlated in every knowledge condition of all three self-conscious scenarios. Hence, we combined these self-conscious emotions and averaged their ratings (reliabilities of these 15 *SCE* scale scores ranged from $\alpha = .791$ to $\alpha = .948$). Similarly, self-reported physical reactions associated with the self-conscious emotions—blush, cover or touch face, hang head down, nervous laughter, sheepish grin, slump shoulders, and stammer[4]—were also highly correlated, so once again we averaged these ratings (reliabilities of these 15 scale scores ranged from $\alpha = .682$ to $\alpha = .850$).[5] Finally, to distinguish self-conscious emotions from general negativity, we subtracted a composite of the ratings of the other negative emotions from the combined self-conscious emotion measure.

2.2. Results

2.2.1. Self-conscious emotion scenarios

Fig. 1 shows the self-conscious emotion ratings (Fig. 1a) and associated physical reactions (Fig. 1b) across knowledge levels in each of the three self-conscious scenarios. Repeated-measures ANOVAs (see Table 1) confirmed that reported self-conscious emotions significantly differed across knowledge levels in the falsifying, teasing, and flatulating scenarios. Planned pair-wise tests across adjacent knowledge levels confirmed that in all three scenarios, self-conscious emotion ratings were higher in the common knowledge condition than in either of the shared knowledge conditions (all $ps < .001$), and higher in each of the shared knowledge than in either of the private knowledge conditions (all $ps < .001$). Repeated-measures ANOVAs also confirmed that the reported physical reactions associated with self-conscious emotions significantly differed across knowledge levels in the falsifying, $F(4, 256) = 23.12$, $p < .001$, $\eta_p^2 = .27$, teasing, $F(4, 308)$

---

[4] Gaze avoidance was not included in these composites, even though it is often associated with the self-conscious emotions, because avoiding or making eye contact can alter the prevailing level of knowledge.
[5] Mean emotion ratings for each knowledge condition and scenario are provided in Supplementary Materials 1.
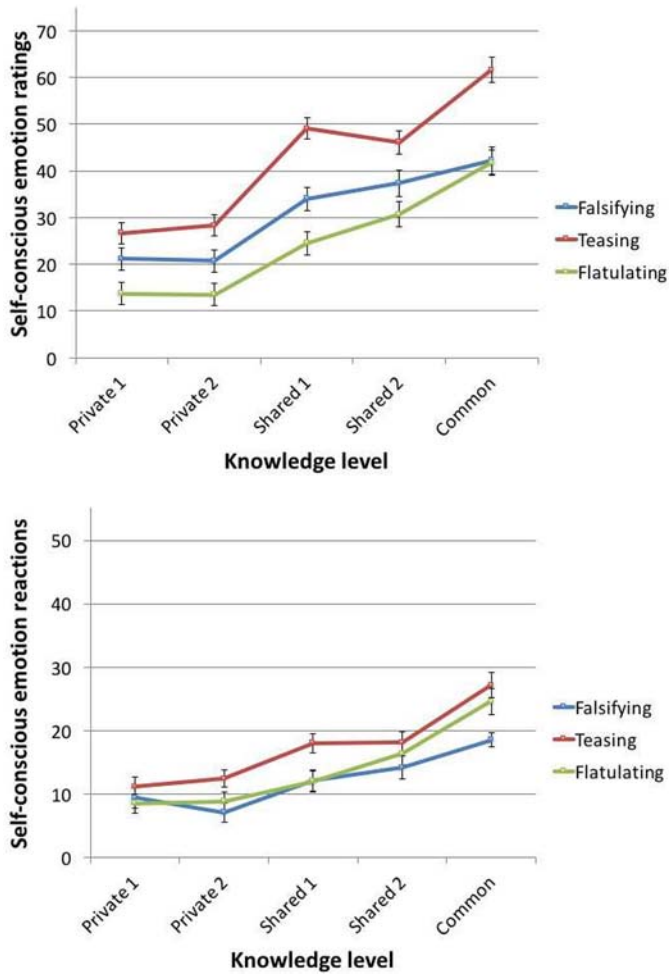
Fig. 1. Ratings of self-conscious emotions (1a, on the top) and likelihood of displaying their physical concomitants (1b, on the bottom) in the three self-conscious emotion scenarios, across the different knowledge levels. Each line in Fig. 1a represents the combined average ratings of embarrassment, shame, and guilt in the respective scenarios. Each line in Fig. 1b represents the combined average ratings of seven physical reactions associated with these emotions in the respective scenarios. Error bars represent standard error.

$= 36.95, p < .001, \eta_p^2 = .32$, and flatulating scenarios, $F(4, 276) = 51.08$, $p < .001, \eta_p^2 = 0.43$. Planned pair-wise tests across adjacent knowledge levels confirmed that in all three scenarios ratings of these reactions were higher in the common than in either of the shared knowledge conditions (all $ps < .001$), and higher in each of the shared than in either of the private knowledge conditions (all $ps < .005$, except $p = .016$ for shared knowledge 1 vs. private knowledge 1 in the falsifying scenario).

We next examine whether sensitivity to knowledge levels is characteristic of negative self-conscious emotions in particular, as the coordination theory predicts, or applies to any negative emotions. We analyzed whether knowledge levels affected participants' ratings of

Table 1
Repeated-measure ANOVA results across all knowledge levels for average self-conscious emotion (SCE) ratings and average basic emotion (BE) ratings.

Scenario		SCE ratings			BE ratings		
		F	df_1, df_2	η_p^2	F	df_1, df_2	η_p^2
SCE scenarios	Teasing	53.16	4, 308	0.41	21.67	4, 308	0.22
	Flatulating	86.65	4, 276	0.56	11.08	4, 276	0.14
	Falsifying	49.02	4, 256	0.43	15.29	4, 256	0.19
BE scenarios	Break up	46.51	4, 304	0.38	203.01	4, 304	0.73
	Sabotage	1.72	4, 280	0.02	86.90	4, 280	0.55

Note: All $ps < 0.001$, except for SCE ratings in the Sabotage scenario where $p = 0.146$.
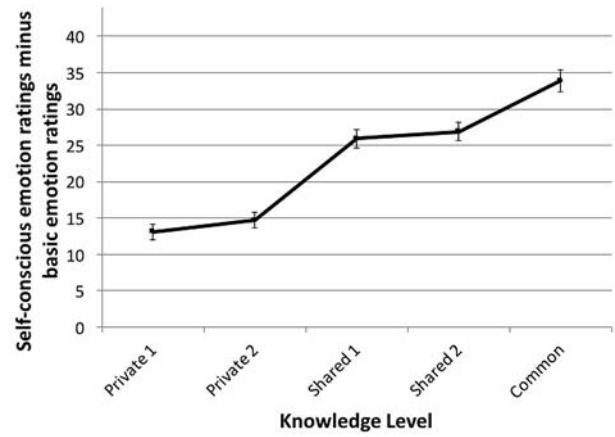
Fig. 2. Average self-conscious emotion ratings minus average basic emotion ratings, collapsed across all three self-conscious emotion scenarios in Experiment 1. Error bars represent standard error.

the emotions of anger, fear, and sadness, which we averaged to form a single index of negative affect. This measure of general negative affect showed a slight increase across knowledge levels in all three scenarios, but critically, the effects of knowledge level were smaller (ranging from $\eta_p^2 = .14$ to $\eta_p^2 = .22$) than for the self-conscious emotions (ranging from $\eta_p^2 = .41$ to $\eta_p^2 = .43$).[6]

We next tested whether knowledge levels shape self-conscious emotions, specifically, after subtracting out general negative affect. We subtracted the negative affect composite (described above) from the self-conscious emotion composite for each participant within each knowledge level, and then collapsed these difference scores across the three scenarios. Fig. 2 shows participants' distinctive self-conscious emotions by knowledge level, and again the pattern supports the coordination hypothesis. Specifically, we find a significant overall difference, $F(4, 848) = 107.21, p < .001, \eta_p^2 = .336$, with common knowledge greater than either of the shared knowledge conditions (both $ps < .001$), and each shared knowledge condition greater than either of the private knowledge conditions (all $ps < .001$).

2.2.2. Basic emotion scenarios

We next analyze whether the non-self-conscious emotions of anger and sadness are sensitive to knowledge levels in the scenarios designed to elicit these emotions. Fig. 3 shows ratings of sadness in the breakup scenario and anger in the sabotage scenario across knowledge levels (Fig. 3a), and the respective physical responses of crying and teeth-clenching (Fig. 3b). Repeated-measures ANOVAs (see Table 1) revealed significant differences across knowledge levels for ratings of sadness in the breakup scenario, and anger in the sabotage scenario, as well as crying in the breakup scenario, $F(4, 304) = 72.33, p < .001, \eta_p^2 = .49$, and teeth-clenching in the sabotage scenario, $F(4, 280) = 34.42, p < .001, \eta_p^2 = .33$.

Post hoc tests showed that these differences across knowledge levels were primarily due to a large difference between responses in the no-knowledge condition and the other four conditions (all $ps < .001$). Sadness ratings in the breakup scenario did not significantly differ across the four conditions in which the participant knew of the event (all $ps > .19$); however, the reported likelihood of crying was significantly lower in the common knowledge condition than in any of the shared or private knowledge conditions (all $ps < .001$). Anger ratings in the sabotage scenario did not differ between the private and shared knowledge conditions (all $ps > .64$); however, anger ratings were significantly

---

[6] See Table 1 for inferential statistics, and Supplementary Materials 1 for descriptive statistics.
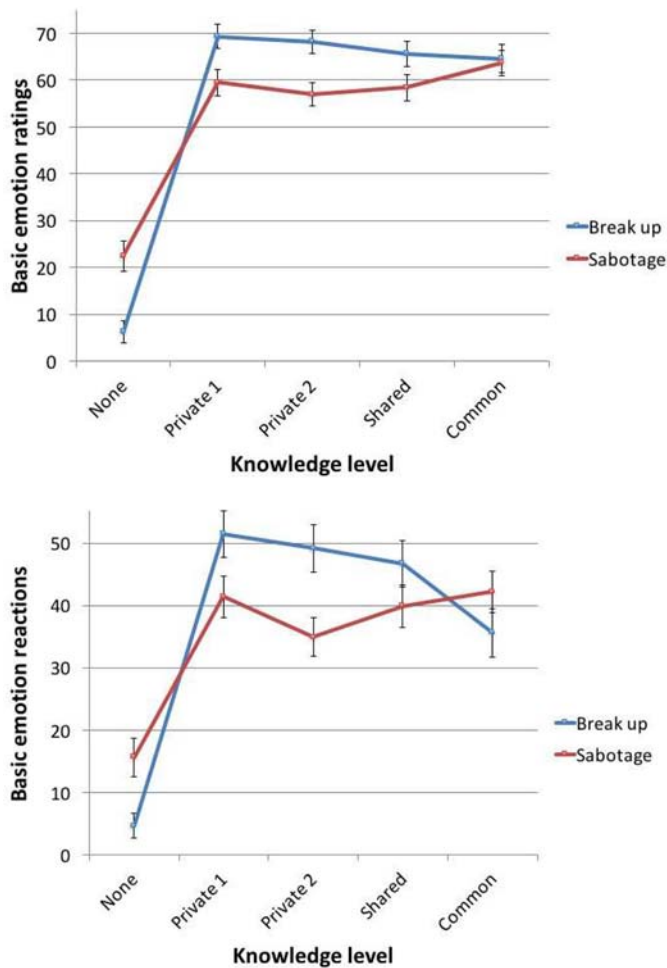
Fig. 3. Ratings of sadness in the break up scenario and anger in the sabotage scenario (3a, on the top), and ratings of the likelihood of crying in the break up scenario and teeth clenching in the sabotage scenario (3b, on the bottom), across the different knowledge levels. Error bars represent standard error.

greater in the common knowledge condition than in the shared knowledge condition ($p = .004$). Unexpectedly, participants reported a lower likelihood of clenching their teeth in the private knowledge 2 condition than in the common knowledge condition ($p = .015$) and the private knowledge 1 condition ($p = .009$).

2.3. Discussion

The results of Experiment 1 replicate the well-known finding that self-conscious emotions are felt more intensely, and are more likely to be expressed in characteristic physical reactions, when a transgression is witnessed by onlookers (shared knowledge), compared to when it is known only privately (private knowledge). More importantly, the results show that people report that they would feel more intense self-conscious emotions when their transgression is common knowledge than when it is shared knowledge. This observation confirms a distinctive prediction of the coordination hypothesis for self-conscious emotions. Moreover, these effects were particular to the self-conscious emotions. Scenarios that elicit anger and sadness showed a different pattern: Anticipated anger and sadness were evoked mainly by privately learning about the triggering event, with little or no increase as knowledge levels increased beyond private knowledge. Hence, people expect that these non-self-conscious emotions are evoked simply by knowing about an eliciting event, whereas the self-conscious emotions track different knowledge levels shared with an audience about the event.

3. Experiment 2

Experiment 2 brings the phenomenon into the lab to test whether embarrassment is intensified by common knowledge of an actual, as opposed to a merely imagined, compromising situation. Conventional wisdom and laboratory research (Hofmann, Moscovitch, & Kim, 2006) both suggest that for most people, singing to an audience is a social challenge with the potential for significant embarrassment, presumably because the singer is making a demand on the audience's attention with the implicit promise to entertain them but with a high probability of failure (particularly since self-consciousness about one's performance itself compromises that performance).

In this experiment, participants sang for a panel of four judges in a separate room through a live two-way video feed and then reported how strongly they felt different emotions. During their performance, participants viewed a karaoke-style video with the lyrics to the song; the lyrics were embedded in a larger video display showing what participants thought was a live feed of the judges in the other room, and which forced them to view the judges for their entire performance (see Fig. 4a and b). In reality, participants viewed a pre-recorded video of a panel of judges rather than live judges. Immediately after their performance, participants rated how they felt for a battery of emotions. The main dependent measures were participants' raw ratings of embarrassment as well as a more refined measure of distinctive embarrassment, in which we subtract general negative affect (average ratings for other negative emotions) from participants' embarrassment ratings.

In a between-subject design, we manipulated two factors related to levels of knowledge. The first manipulation was that the participant was told either: (1) the judges knew there was a live two-way video feed, and knew that the participant was aware they knew this (common knowledge condition); or, (2) the judges falsely believed they were

Fig. 4. Representative screenshots of the videos participants saw in the eye contact (4a, on the top), and no eye contact conditions (4b, on the bottom) in Experiment 2. The black screen with words in the lower left corner was an embedded video that provided lyrics to the karaoke song.

watching a one-way video feed and thought that the participant was unaware of their presence (shared knowledge condition). The second manipulation was that the judges in the video were either looking straight into the camera (eye contact condition; Fig. 4a) or they were looking off to the side (no eye contact condition; Fig. 4b). These manipulations were designed to vary participants' beliefs about common knowledge and eye contact as a psychophysical cue for common knowledge.

The coordination theory predicts greater embarrassment when a participant thinks that their bad singing is common knowledge. In the shared knowledge condition, a participant is singing badly and knows that strangers are watching, but at least the strangers think he is singing only to himself, and moreover, these strangers could still pretend they didn't see what happened to help the participant save face. However, in the common knowledge condition, a participant cannot take solace in these considerations: The fact that the judges saw his bad singing in a two-way video and that this was common knowledge means that neither party can pretend ignorance of the other. The participant's bad signing is an unavoidable and conspicuous fact before them and the judges.

Hence, by comparing shared knowledge and common knowledge, we can test whether the self-conscious emotion of embarrassment is sensitive to distinct levels of knowledge. If people feel embarrassed mainly in response to what others know about their blunders, then they should feel similarly embarrassed when the judges see their bad singing, whether or not the judges know the participant knows they are watching. But, if people feel extra embarrassment when their blunders are common knowledge, then participants will be more embarrassed when everyone is aware that the judges saw their bad singing.

This experiment thus tests the coordination theory's prediction about how embarrassment will be affected by different knowledge levels (shared vs. common knowledge), and by a potential psychophysical cue of common knowledge (eye contact vs. no eye contact) in a 2 × 2 factorial between-subjects design.

3.1. Method

3.1.1. Participants

Based on a power analysis for detecting medium effect sizes, we recruited 126 participants (aged 18–75) from the Harvard University study pool, who received $10 or course credit for participation. Because participant suspicion in experiments involving deception is a specific concern, this was assessed using a structured funnel debriefing, and participants that exhibited high levels of suspicion ($n = 10$), or that didn't actually perform the karaoke song ($n = 3$) were excluded. We had a final sample of $N = 113$ (57% female, $M_{age} = 28.1$, $SD_{age} = 13.3$).[7]

3.1.2. Procedure

Participants signed up for a study described as an experiment on how anonymity affects judges' evaluations of someone's performance, and which stated that they would be assigned either to perform a karaoke song or to judge someone else's performance. Upon arrival, participants were told that we were studying how the relationship between judges and performers affects judges' evaluations—"an important unanswered question in the age of reality TV"—and that in each study we recruit a panel of judges and a performer. All participants were then informed that they had been assigned to the performer role, and that four other participants assigned to the role of judges were already situated in a room across the hallway. Participants were then told that they would sing in a karaoke performance that the judges would observe

over a live two-way video feed, and that this two-room set up was necessary to ensure that the environment was the same when judges believed they were anonymous and when they did not.

Participants assigned to the common knowledge condition were then told: "You have been selected for the condition in which the judges know that you are aware of their presence. You will be able to see them on the two-way video feed during your performance, and they know that you can see them through the two-way video feed. In other words, both you and the judges know that you are watching each other over a two-way video feed during your performance." Thus, in this common knowledge condition, participants believed that everyone knew that everyone knew (and so on) that the judges were watching the participant (and vice versa) over a live two-way video feed.

Participants assigned to the shared knowledge condition were instead told: "You have been selected for the anonymous condition, in which the judges think that you are unaware of their presence. However, in order to keep the performances the same across conditions, we actually always tell the performers about the judges. In addition, you will be able to see them on the two-way video feed during your performance, but they don't know this. In other words, even though you can see the judges over this feed, they think that you don't know they are there, and that they are simply watching a one-way video feed of your performance." Thus, in this shared knowledge condition, participants believed that the judges thought they were watching a one-way video feed and that the participant was unaware of their presence (even though the participant actually knew they were there and could see them over the live two-way feed).

Participants in both conditions were then told that after the song ended, the video feed would cut off, and they would be given a short survey.[8]

Participants were then escorted into a small private room with a large computer monitor that had a built-in video camera, speakers, and a microphone for them to sing into. The researcher set the computer up in front of the participant, and then told the participant that she had to go set up the video feed in the judges' room. Participants were told that as soon as the researcher got this set up, they would see a video feed of the judges as well as a karaoke video with the song lyrics, and were instructed to begin their performance as soon as the song started. After 30 s, a video appeared on the computer screen with the researcher's head right in front of the camera (as though she was just turning on the video camera in the other room). Immediately thereafter, Adele's *Rolling in the Deep*, a well-known pop song with a soaring chorus, began to play, and participants saw a large video image of the judges in the other room, with a small karaoke feed embedded in the lower left corner that displayed the lyrics to the song (see Fig. 4). In the *eye contact* condition, the video of the judges showed them looking directly into the camera (Fig. 4a); in the *no eye contact* condition, the video showed them looking askance, as though they were watching on a screen a few feet to the side of the camera (Fig. 4b).

After the song finished, the video shut off and a survey popped up on the computer screen, asking participants to rate how strongly they experienced different emotions on a 1–5 scale in a slightly modified Positive and Negative Affect Schedule (PANAS). The PANAS includes 10 items comprising a negative affect scale and 10 items comprising a positive affect scale (see Watson et al., 1988), and was used to disguise our interest in embarrassment by embedding it in a larger battery of emotion ratings. However, the PANAS does not specifically ask about embarrassment, and so one of the positive affect items was replaced with *embarrassed*.

Participants then answered a few questions about their experience with karaoke and provided basic demographic information. They then exited the room, where they met the researcher, and were given a funnel debriefing. The debriefing began with vague questions about what

---

[7] The questions used in the funnel debriefing are provided in Supplementary Materials 4, and descriptive statistics and sample sizes are provided in Supplementary Materials 5. All exclusion criteria were specified prior to data collection, and all exclusion decisions were made prior to any statistical analysis of the data.

[8] The complete instructions are supplied in Supplementary Materials 3.
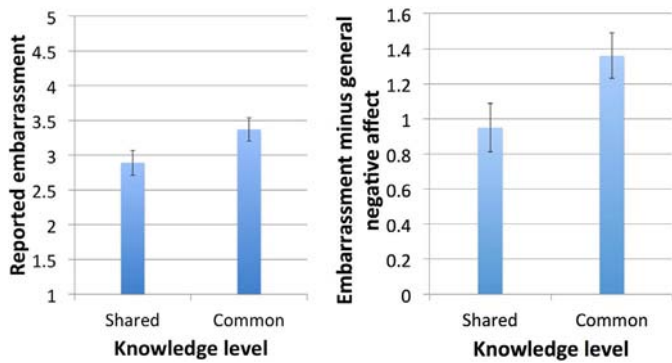
Fig. 5. Average reported embarrassment (5a, on the left), and average reported embarrassment with general negative affect subtracted out (5b, on the right), in the shared and common knowledge conditions in Experiment 2. Error bars represent standard error.

participants thought the study was about, followed by increasingly specific questions about whether they thought everything had been clearly explained to them and whether they were suspicious during the study, concluding with a direct question about whether they believed the judges were real.[9] Participants were then fully debriefed and paid.

3.2. Results and discussion

Fig. 5a shows the results. A 2 (shared vs. common knowledge) × 2 (eye contact vs. no eye contact) ANOVA revealed a marginally significant effect of knowledge level on ratings of embarrassment, $F(3, 109)$ = 3.83, p = .053, η_p^2 = .03 (see Fig. 5a), no effect of eye contact, $F(3, 109)$ = .96, p = .328, η_p^2 = .01, and no interaction effect, $F(3, 109)$ = .15, p = 0.703, η_p^2 = .00. Specifically, participants showed a trend of greater embarrassment in the common knowledge than shared knowledge condition, consistent with the coordination hypothesis, although we note that this trend had marginal significance.

Next, as in Experiment 1, we examined participants' distinctive embarrassment by subtracting other negative emotions—in this case, an average of the 10 items of the negative affect scale from the PANAS (α = .891). A 2 × 2 ANOVA revealed a significant effect of knowledge level on embarrassment minus negative affect, $F(3, 109)$ = 4.77, p = .031, η_p^2 = .04 (see Fig. 5b), no effect of eye contact, $F(3, 109)$ = 1.82, p = .181, η_p^2 = .02, and no interaction effect, $F(3, 109)$ = .16, p = .691, η_p^2 = .00. Specifically, participants showed greater distinctive embarrassment when their karaoke performance was common knowledge than when it was shared knowledge.

Overall, these results provide some initial support for the prediction that people feel more embarrassed when their performance is common knowledge than when it is shared knowledge. The effect was marginally significant for raw embarrassment ratings and reached significance for distinctive embarrassment. In contrast, we found no evidence that eye contact, a potential common-knowledge generator, itself increases embarrassment.

4. General discussion

Overall, these experiments provide two lines of converging evidence that acknowledging a transgression with an audience evokes more intense feelings of guilt, embarrassment, and shame than simply knowing they know about it. Drawing on insights from game theory, the coordination theory proposes that people feel more embarrassed, guilty, or

ashamed when a transgression is acknowledged, or otherwise becomes public, for two reasons: (1) An actor's failure to atone for a transgression conveys more damaging information when it is shared knowledge compared to when an audience simply knows about it; and, (2) Human relationships are a kind of coordination game, which makes a transgression more disruptive when it is common knowledge compared to shared knowledge. This coordination theory makes the novel prediction that the self-conscious emotions will be elicited more strongly by common knowledge of a transgression than by shared knowledge.

In Experiment 1, a typical participant reported that he would feel more embarrassed, guilty, and ashamed in hypothetical scenarios when he and an observer had common knowledge of the offense than when he and the observer had only shared knowledge. We observed these effects for participants' raw ratings of self-conscious emotions and also for a refined measure of distinctive self-conscious emotions, subtracting general negative affect. Moreover, the effects of shared vs. common knowledge were more pronounced for self-conscious emotions, than for two other negative emotions, sadness and anger.

One empirical complication in Experiment 1, is that participants also reported somewhat greater anger when an aggressive act was common knowledge between aggressor and victim rather than merely shared knowledge. This observation might hint that even some so-called basic emotions could also help coordinate relationships, in addition to other core functions. For instance, anger may involve elements of coordination as two parties renegotiate their joint expectations of the deference one owes the other (see Reed, DeScioli, & Pinker, 2014; Sell, Tooby, & Cosmides, 2009).

In Experiment 2, a typical participant reported feeling more embarrassed by a karaoke performance in front of a panel of judges when he believed the judges knew he was aware of their presence than when he believed the judges thought they were anonymous. In this case, however, participants' raw ratings of embarrassment reached only marginal significance. But, when we examined a refined measure of distinctive embarrassment, after subtracting negative affect, we found significantly greater values for common knowledge than shared knowledge. Although these results are somewhat mixed, we suggest that overall they converge with Experiment 1 to point to more intense self-conscious emotions under common knowledge than shared knowledge. In contrast, we did not find effects of eye contact, which we theorized would be a potent cue of common knowledge; we further discuss this observation below.

Hence, these findings suggest a key addendum to previous reputation theories for self-conscious emotions: Realizing that other people know about one's offense can be uncomfortable, but realizing that they know that one knows that they know (and so on) can be agonizing. We suggest that this is because self-conscious emotions help maintain reputation in addition to maintaining relationships by coordinating people's beliefs about their relationship status.

4.1. Social signals and atonement

Someone who breaches social expectations without acknowledging the transgression is merely under a cloud of suspicion, and the possibility remains open that he will establish his innocence, compensate the victim, or mend his ways—that is, that he at least wants to play by the rules and is sufficiently competent to do so. In contrast, someone who acknowledges the breach yet proceeds without apology indicates that he does not consider himself bound by the community's rules (or is unable to understand or abide by them) and sees little value in belonging to that community. Furthermore, as information about the transgression leaks and spreads, the credibility of any eventual apology diminishes (Sznycer et al., 2015). The acute discomfort triggered by shared or common knowledge of a transgression, then, serves to motivate the transgressor to atone for his actions, make such atonements more credible, and thereby protect his threatened standing in the community.

4.2. Relationships as coordination games

The analysis presented in this paper holds that social relationships can be characterized as coordination games. Two individuals can benefit if they both tacitly agree to be friends, lovers, helpful neighbors, trading partners, boss and employee, or polite strangers. These social agreements allow both partners to share expectations about resources, roles, and acceptable behavior without constant negotiation, misunderstandings, and conflict (Fiske, 1992; Haslam, 2004; Lee & Pinker, 2010; Pinker, 2007). But if two people assume different relationship models (lover versus supervisee, friend versus boss, customer versus family member), conflicts can arise as one party claims resources or perquisites that the other is unwilling to cede, resulting in feelings of impoliteness, imposition, insubordination, bossiness, sexual harassment, exploitation, or inappropriate familiarity or formality. To effectively manage different types of relationships, the self-conscious emotions help solve these coordination problems by establishing and repairing common knowledge of relationship types.

Previous work on knowledge levels and coordination has primarily focused on how common knowledge can help social partners achieve mutual benefits by coordinating on new opportunities (e.g., Chwe, 2001; Schelling, 1960; Thomas et al., 2014). However, common knowledge can also generate mutual losses by triggering a runaway cascade in which established partners second-guess each other's future intentions, causing coordination to unravel.

Common knowledge can thus both facilitate coordination and dismantle it. It can help people build new relationships but can also destroy existing ones. This is because coordination relies on an unstable foundation of mutual expectations. When trying to coordinate with a partner, an individual should do what he expects his partner will do, which in turn depends on what he expects the partner to expect him to do, and on what he expects she expects he expects her to do, and so on. Common knowledge can facilitate coordination by focusing partners' expectations on the same option, so that these interdependent expectations line up. Common knowledge can also destroy coordination by focusing partners' attention on an event that suggests their interests may no longer be aligned, which can make expectations about future behavior unpredictable again. This shift in expectations can present cooperative partners with the same dilemma they faced when they first established their relationship, and requires them to take action to realign expectations if they want to reestablish the status quo.

Shared knowledge, in contrast, is less potent in both respects, because without common knowledge, the two partners cannot be assured that their expectations of future behavior are aligned. Without a clear alignment of expectations, shared knowledge can make it difficult to coordinate and establish a new relationship. However, this same lack of alignment also means that shared knowledge of a transgression can leave cooperative partners with enough wiggle room to avoid the potentially painful and costly process of renegotiating the terms of their relationship. Put simply, this allows established cooperative partners to plausibly pretend that nothing ever happened and more easily resume the status quo if they want to.

4.3. Third-party coordination and self-conscious emotions

In addition to the coordination problem within a relationship, there are also coordination problems for third parties outside of a relationship. If a third-party observer accuses someone of wrongdoing or incompetence, then this observer puts him or herself in harm's way. The accused person could retaliate with counteraccusations and damaging attacks. However, these costs are reduced if other observers do the same because the violator's retaliation is now divided among a group of accusers. Hence, third parties who want to chastise, exclude, or punish a violator are posed with a coordination problem: If an individual confronts the violator alone, the costs of retaliation could be steep, but

if multiple observers coordinate their accusations, then the threat of retaliation is minimized.

As in other coordination problems, third parties are better equipped to synchronize their condemnation if they have common knowledge about the violator's infraction (DeScioli & Kurzban, 2013). Thus, punishment is less costly and therefore more likely when a transgression is common knowledge among observers compared to if it is only shared knowledge.

From the transgressor's perspective, this means that common knowledge of his offense among multiple observers poses a greater threat of punishment, and so a greater need to atone for his violation. The self-conscious emotions can better account for threat of punishment by making the motive to apologize even greater when a violation is common knowledge among observers. That is, the self-conscious emotions might be designed to help people manage not only their own dyadic relationships after a transgression, but also the broader social threats that come from multiple third parties who could become aware of a violator's indiscretions. If so, then self-conscious emotions might have also evolved to track knowledge levels among third parties, in addition to between oneself and a social partner.

This extension of the coordination theory of self-conscious emotions offers a number of testable predictions for future research. Most obvious, it predicts that self-conscious emotions will be more intense when a violation is common knowledge among observers—for example, if the transgression occurs in public and is visible to many people (who can also see each other observing it), or if the violation is recorded in photos or videos than can be widely observed in public forums such as Twitter or YouTube. A second prediction is that people's self-conscious emotions will be sensitive to cues indicating that an increasing number of observers are acquiring common knowledge of the offense, such as overhearing others gossip about the transgression. A third prediction is that self-conscious emotions will motivate people to try to prevent their offense from becoming common knowledge, such as coercing others not to speak of the offense, even if all members of the audience already privately knew about it.

4.4. Common knowledge and the display of self-conscious emotions

If the self-conscious emotions evolved according to the logic of coordination games, this might help explain another distinctive feature: They trigger not only localized facial displays, as the basic emotions do, but also conspicuous autonomic responses and dynamic body postures that alter the visibility of one's face (Keltner, 1995; Keltner & Buswell, 1997; Lewis, 2000; Tracy & Matsumoto, 2008; Tracy & Robins, 2004). Shame and embarrassment can cause blushing, slumping the shoulders, hanging the head, avoiding others' gaze, and covering the face. Pride, a positive self-conscious emotion, is associated with puffing the chest and raising the arms and head. We suggest these displays might have evolved in part to generate or avoid common knowledge.

Much of the discomfort of blushing comes from the fact that it is both displayed as redness on the outside and felt as heat and tingling on the inside, so the blusher knows that onlookers know that he knows that they know he is blushing, and so on (see Thomas et al., 2014). The downcast bodily postures accompanying shame may work on two levels. By averting eye contact, the transgressor is seemingly avoiding common knowledge of the transgression. Yet unlike a furtive glance away of a transgressor who is still hoping to evade common knowledge, the conspicuous downcast posture of an ashamed person signals that he is aware of the need to avoid common knowledge while no longer actually trying to avoid it. Not only can an observer see that a slumping transgressor's emotional state has changed, but the transgressor's own view of the observers has been obscured or reoriented away, and crucially, the observer can see that, and see that the transgressor must be aware that the observer can see it, and so on. (As Goffman, 1959, 1978, and Brown & Levinson, 1987, point out, it's no coincidence that we say that a person who has committed a

commonly known social infraction has "lost face" and that a person attempting to mitigate the harm is trying to "save face.") A downcast individual can thus make it common knowledge that he already recognizes the need to change his behavior, which can allow all parties to resume the status quo without having to engage in unnecessary conflict. Thus, the physical expression of self-conscious emotions is consistent with the theory that these emotions, to a greater extent than other emotions, involve recursive mental state reasoning in the service of strategies for negotiating the kinds of coordination games that permeate human social life.

One empirical wrinkle in this explanation is that in Experiment 2, eye contact, which should theoretically be a powerful common knowledge generator (each person sees the other looking back), did not in fact lead to greater embarrassment. This observation is surprising given the role of eye contact in social life in general, and specifically the phenomenon of gaze aversion in response to embarrassment. The finding may mean that direct eye contact does not affect embarrassment, or does not lead to inferences of common knowledge. However, we note that the current manipulation of eye contact was relatively subtle and unconventional; it occurred via a video feed in which the judges' eyes were oriented directly toward or at a slight angle relative to the participant. Moreover, the judges were actually taped beforehand, so their eyes could not dynamically respond to the participant's actions, which may have appeared as staring blankly rather than attending to the participant. Future research can examine the effects of eye contact under more naturalistic settings to further test how it may relate to embarrassment and inferences about common knowledge.

5. Conclusion

We close by suggesting that common knowledge has far-reaching implications for psychology. The categorical difference between shared and common knowledge has long been recognized by game theorists (Rubinstein, 1989; Schelling, 1960), economists (Chwe, 2001; Geanakoplos, 1992), philosophers (Lewis, 1969), linguists (Clark, 1996; Clark & Marshall, 1981; Smith, 1982), sociologists (Adut, 2008), anthropologists and political scientists (Cronk & Leech, 2012), legal theorists (McAdams & Nadler, 2005), and computer scientists (Alberucci & Jäger, 2005; Halpern & Moses, 1990). Recent research has demonstrated that this distinction also shapes the psychology of cooperation (Thomas et al., 2014), helping (Thomas et al., 2016), and communication (Lee & Pinker, 2010). The present experiments add self-conscious emotions to the set of human psychological systems that are sensitive to the categorical differences between private, shared, and common knowledge.

Acknowledgements

We thank Tiana Press for her assistance in running experiments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2017.12.001.

References

Adut, A. (2008). *On scandal: Moral disturbances in society, politics, and art.* New York, NY: Cambridge University Press.

Alberucci, L., & Jäger, G. (2005). About cut elimination for logics of common knowledge. *Ann. Pure Appl. Logic, 133*(1–3), 73–99. https://doi.org/10.1016/j.apal.2004.10.004.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323.

Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D., & Knight, R. T. (2003). The regulatory function of self-conscious emotions: Insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology, 85*(4), 594–604. https://doi.org/10.1037/0022-3514.85.4.594.

Bosch, J. A., de Gues, E. J. C., Carroll, D., Goedhart, A. D., Anane, L. A., Veldhuizen, J. J., ... Edwards, K. M. (2009). A general enhancement of autonomic and cortisol responses during social evaluative threat. *Psychosomatic Medicine, 71*, 877–885. https://doi.org/10.1097/PSY.0b013e3181baef05.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage.* New York, NY: Cambridge University press.

Chakroff, A., Thomas, K. A., Haque, O. S., & Young, L. (2015). An indecent proposal: The dual functions of indirect speech. *Cognitive Science, 39*(1), 199–211. https://doi.org/10.1111/cogs.12145.

Chwe, M. S. (2001). *Rational ritual: Culture, coordination, and common knowledge.* Princeton, NJ: Princeton University Press.

Clark, H. H. (1996). *Using language.* Cambridge, UK: Cambridge University Press.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, UK: Cambridge University Press.

Cronk, L., & Leech, B. L. (2012). *Meeting at grand central: Understanding the social and evolutionary roots of cooperation.* Princeton, NJ: Princeton University Press.

Dalkiran, N. A., Hoffman, M., Paturi, R., Ricketts, D., & Vattani, A. (2012). Common knowledge and state-dependent equilibria. *Proceedings of symposium on algorithmic game theory* (pp. 84–95). https://doi.org/10.1007/978-3-642-33996-7_8.

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin, 139*(2), 477–496. https://doi.org/10.1037/a0029065.

Dijk, C., de Jong, P. J., & Peters, M. L. (2009). The remedial value of blushing in the context of transgressions and mishaps. *Emotion, 9*(2), 287–291. https://doi.org/10.1037/a0015081.

Dijk, C., Koenig, B., Ketelaar, T., & de Jong, P. J. (2011). Saved by the blush: Being trusted despite defecting. *Emotion, 11*(2), 313–319. https://doi.org/10.1037/a0022774.

Ekman, P. (1999). Basic emotions. In T. Dalgleish, & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). West Sussex, UK: John Wiley & Sons Ltd.

Ekman, P. (2003). *Emotions revealed.* New York, NY: Times Books.

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review, 99*(4), 689–723. https://doi.org/10.1037/0033-295X.99.4.689.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions.* New York, NY: W. W. Norton.

Geanakoplos, J. (1992). Common knowledge. *The Journal of Economic Perspectives, 6*(4), 53–82. https://doi.org/10.1257/jep.6.4.53.

Goffman, E. (1959). *The presentation of self in everyday life.* New York, NY: Anchor Books.

Goffman, E. (1978). Response cries. *Language, 54*(4), 787–815. https://doi.org/10.2307/413235.

Gruenewald, T. L., Kemeny, M. E., Aziz, N., & Fahey, J. L. (2004). Acute threat to the social self: Shame, social self-esteem, and cortisol activity. *Psychosomatic Medicine, 66*, 915–924. https://doi.org/10.1097/01.psy.0000143639.61693.ef.

Halpern, J. Y., & Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM, 37*(3), 549–587. https://doi.org/10.1145/79147.79161.

Haslam, N. (2004). Research on the relational models: An overview. In N. Haslam (Ed.), *Relational models theory: A contemporary overview* (pp. 27–57). Mahwah, NJ: Erlbaum.

Hofmann, S. G., Moscovitch, D. A., & Kim, H. (2006). Autonomic correlates of social anxiety and embarrassment in shy and non-shy individuals. *International Journal of Psychophysiology, 61*(2), 134–142. https://doi.org/10.1016/j.ijpsycho.2005.09.003.

Holmes, E. A., Arntz, A., & Smucker, M. R. (2007). Imagery rescripting in cognitive behavior therapy: Images, treatment techniques and outcomes. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 297–305. https://doi.org/10.1016/j.jbtep.2007.10.007.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68*(3), 441–454. https://doi.org/10.1037/0022-3514.68.3.441.

Keltner, D., & Buswell, B. N. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion, 10*(2), 155–171. https://doi.org/10.1080/026999396380312.

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122*(3), 250–270. https://doi.org/10.1037/0033-2909.122.3.250.

Ketelaar, T., & Au, W. T. (2003). The effects of feelings of guilt on the behavior of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition and Emotion, 17*(3), 429–453. https://doi.org/10.1080/02699930143000662.

Lee, J. J., & Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review, 117*(3), 785–807. https://doi.org/10.1037/a0019688.

Lewis, D. (1969). *Convention: A philosophical study.* Cambridge, MA: Harvard University Press.

Lewis, M. (2000). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 623–636) (2nd ed.). New York: Guilford.

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science, 3*(3), 173–192. https://doi.org/10.1111/j.1745-6924.2008.0073.x.

McAdams, R. H., & Nadler, J. (2005). Testing the focal point theory of legal compliance: The effect of third-party expression in an experimental hawk/dove game. *Journal of Empirical Legal Studies, 2*(1), 87–123. https://doi.org/10.1111/j.1740-1461.2005.00032.x.

McCullough, M. E. (2008). *Beyond revenge: The evolution of the forgiveness instinct.* San Francisco, CA: Jossey-Bass.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2010). Evolved mechanisms for revenge and forgiveness. In P. R. Shaver, & M. Mikulincer (Eds.), *Understanding and reducing aggression, violence, and their consequences* (pp. 221–239). Washington, DC: American Psychological Association.

Miller, R. S. (1995). On the nature of embarrassability: Shyness, social evaluation, and social skill. *Journal of Personality*, 63(2), 315–339. https://doi.org/10.1111/j.1467-6494.1995.tb00812.x.

Modigliani, A. (1971). Embarrassment, facework, and eye contact: Testing a theory of embarrassment. *Journal of Personality and Social Psychology*, 17(1), 15–24. https://doi.org/10.1037/h0030460.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature.* New York, NY: Viking.

Pinker, S. (2011). *The better angels of our nature: Why violence has declined.* New York, NY: Viking.

Pinker, S. (2014). *The sense of style: The thinking person's guide to writing in the 21st century.* New York, NY: Penguin.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, 105(3), 833–838. https://doi.org/10.1073/pnas.0707192105.

Reed, L. I., DeScioli, P., & Pinker, S. (2014). The commitment function of angry facial expressions. *Psychological Science*, 25(8), 1511–1517. https://doi.org/10.1177/0956797614531027.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2.

Rubin, D. C., & Talarico, J. M. (2009). A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words. *Memory*, 17(8), 802–808. https://doi.org/10.1080/09658210903130764.

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge.". *The American Economic Review*, 79(3), 385–391 Retrieved from http://www.jstor.org/stable/1806851.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 76(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805.

Schelling, T. C. (1960). *The strategy of conflict.* Cambridge, MA: Harvard University Press.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences*, 106(35), 15073–15078 (doi:10.1073pnas.0904312106).

Skyrms, B. (2004). *The stag hunt and the evolution of social structure.* Cambridge, UK: Cambridge University Press.

Smith, N. (1982). *Mutual knowledge.* Orlando, FL: Academic Press.

Smith, R. H., Webster, J. M., Parrot, W. G., & 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. https://doi.org/10.1037//0022-3514.83.1.138.

Steckler, C. M., & Tracy, J. L. (2014). The emotional underpinnings of social status. In J. T. Cheng, J. L. Tracy, & C. Anderson (Eds.), *The psychology of social status* (pp. 201–224). New York, NY: Springer.

Sznycer, D., Schniter, E., Tooby, J., & Cosmides, L. (2015). Regulatory adaptations for delivering information: The case of confession. *Evolution and Human Behavior*, 36, 44–51. https://doi.org/10.1016/j.evolhumbehav.2014.08.008.

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* (Advance online publication) https://doi.org/10.1073/pnas.1514699113.

Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology*, 20(6), 1256–1269. https://doi.org/10.1037/0022-3514.70.6.1256.

Tangney, J. P., & Tracy, J. L. (2012). Self-conscious emotions. In M. Leary, & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 446–478) (2nd ed.). New York, NY: Guilford.

Thomas, K. A., De Freitas, J., DeScioli, P., & Pinker, S. (2016). Recursive mentalizing and common knowledge in the bystander effect. *Journal of Experimental Psychology: General* (Advance online publication) https://doi.org/10.1037/xge0000153.

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, 107(4), 657–676. https://doi.org/10.1037/a0037037.

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 (Eds.), *Evolution of social behavior patterns in primates and man. Proceedings of the British Academy, 88*. (pp. 119–143) Retrieved from http://www.cep.ucsb.edu/papers/friendship.pdf.

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences*, 105(33), 11655–11660. https://doi.org/10.1073/pnas.0802686105.

Tracy, J. L., & Robins, R. W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory*, 15(2), 103–125. https://doi.org/10.1207/s15327965pli1502_01.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46(1), 35–57 Retrieved from http://www.jstor.org/stable/2822435.

Utikal, V. (2012). A fault confessed is half redressed—Confessions and punishment. *Journal of Economic Behavior & Organization*, 81(1), 314–327. https://doi.org/10.1016/j.jebo.2011.11.001.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063.

Weiner, B., Graham, S., Peter, O., & Zmuidinas, M. (1991). Public confession and forgiveness. *Journal of Personality*, 59(2), 281–312. https://doi.org/10.1111/j.1467-6494.1991.tb00777.x.

Willer, R., Kuwabara, K., & Macy, M. W. (2009). The false enforcement of unpopular norms. *American Journal of Sociology*, 115(2), 451–490. https://doi.org/10.1086/599250.

Wolf, S. T., Cohen, T. R., Panter, A. T., & Insko, C. A. (2010). Shame proneness and guilt proneness: Toward the further understanding of reactions to public and private transgressions. *Self and Identity*, 9, 337–362. https://doi.org/10.1080/15298860903106843.