The Commitment Function of Angry Facial Expressions

Citation

Published Version
doi:10.1177/0956797614531027

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:23597724

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
The Commitment Function of Angry Facial Expressions
Lawrence Ian Reed
Harvard University
Peter DeScioli
Stony Brook University
Steven A. Pinker
Harvard University

Author Note
Lawrence Ian Reed, Department of Psychology, Harvard University; Peter DeScioli, Department of Political Science, Stony Brook University; Steven A. Pinker, Department of Psychology, Harvard University.

Correspondence concerning this article should be addressed to Lawrence Ian Reed, Department of Psychology, Harvard University, Cambridge, MA, 02138. E-mail: lawrenceianreed@fas.harvard.edu
Abstract

What are facial expressions of emotion for? We test the hypothesis that some expressions serve as honest signals of subjective commitments, in particular, that angry faces increase the effectiveness of threats. In an ultimatum game, proposers decided how much money to offer a responder while seeing a film clip depicting an angry or a neutral facial expression, together with a written threat that was either inherently credible (a 50-50 split) or less credible (a demand for 70% of the money). Proposers were more likely to accede to the less credible threats when they were accompanied by an angry than a neutral expression, but were unaffected by the expression when dealing with credible threats. This supports the hypothesis that angry expressions are honest signals that enhance the credibility of threats.

*Keywords:* facial expressions, emotions, threat, punishment
Angry expressions are among the most common signals in human social life. An unruly child, a colleague late for a meeting, or a business partner taking more than his or her fair share is likely to elicit an unmistakable glower composed of slanted brows, glaring eyes, and tight lips. An angry face can escalate hostilities and lead to destructive behavior, including violence. Intuitively, people sense that angry expressions may also keep the peace by conveying social expectations and personal boundaries: the threat of aggression can avoid actual aggression. This intuition, however, is theoretically incomplete and empirically unsupported. Here we test a particular version: that angry expressions function as credible threats in bargaining.

The intuitively obvious theory of expressions is that they signal to conspecifics current emotional states (Izard, 1971) or action tendencies (Fridlund, 1994; Kraut & Jonston, 1979). For example, expressions of anger communicate a signaler’s intention to approach a receiver aggressively (van Honk & Schutter, 2005; Yik & Russell, 1999; Adams, Ambady, Macrae, & Kleck, 2006), whereas expressions of happiness signal the absence of threat (Ramachandran, 1998) and cooperative intent (Reed, Zeglen, & Schmidt, 2012).

The problem with a simple signaling theory is that it is not obvious why it is adaptive for an organism to telegraph its internal state or impending action, nor why a perceiver should take such signals at face value. Darwin (1998) argued that expressions may originally have had a noncommunicative adaptive function, namely preparing the organism to respond to environmentally recurrent stimuli by regulating sensory intake. Widened eyes in expressions of fear, for example, increase peripheral vision to identify potential threats (Susskind, Lee, Cusi, Grabski, & Anderson, 2008). Similarly, the constriction of the nose and mouth in disgust expressions functions to expel and prevent internal exposure to noxious stimuli (Chapman, Kim, Susskind, & Anderson, 2009). Darwin argued that even these functions may no longer apply in modern humans: expressions today are evolutionary vestiges, automatized through Lamarckian use and disuse. Other features of expressions merely represent an overflow of energy from an aroused nervous system (Darwin, 1998; Fridlund, 1992).

A breakthrough in specifying a communicative function for facial expressions comes from game theory: the hypothesis that expressions communicate an individual’s commitments to carry out threats or promises in bargaining situations (Schelling, 1960; Hirshleifer, 1987). A commitment ties one’s hands in order to make it difficult to renege on the threat or promise. The goal of commitment is to influence the partner’s choices by leaving them the last opportunity to decide the outcome (Schelling, 1960). This lack of freedom of choice can, paradoxically, improve one’s bargaining position.

An inherent problem in making a commitment is how to signal it credibly. Many commitments are enforced by third parties such as legal systems or constraints in the situation such as burning bridges. Alternatively, a subjective commitment is enforced by the individual’s internal emotional state (Pinker, 1997; Frank R., 1988; Nesse, 2001): the actor is in the throes of an emotion that makes it difficult to renege, even if carrying out the threat or promise is costly. Because this emotional enforcement is internal and unobservable, subjective commitments must be credibly displayed to be effective. This supplies the elusive communicative function for facial expressions.

Two sets of findings support the hypothesis that expressions credibly signal subjective commitments. First, expressions are related to emotional experiences (Rosenberg & Ekman, 2005) and convey information about intentions and action tendencies (Fridlund, 1994; Kraut & Jonston, 1979; Yik & Russell, 1999). Second, expressions are associated with involuntary neurological mechanisms (Rinn, 1984) and are difficult to fake (Ekman, Roper, & Hager, 1980). This association between motivational and expressive systems could allow expressions to be predictive of future behavior and hence credible signals.
A major problem for the credible signaling theory, however, is that individuals can gain advantages through deceptive signaling (Dawkins & Krebs, 1978), fooling perceivers into believing a threat or promise but then reneging if the costs are too high. If actors can dissociate facial displays from underlying psychological states, then they could gain a strategic advantage by deploying these signals deceptively (“bluffing”). In this case, facial expressions would be “cheap talk” and perceivers would evolve or learn to ignore the signals. Only empirical tests can establish whether humans in fact treat emotional expressions as credible signals of bargaining resolve. We report two such tests, focused on expressions of anger.

Anger is an aversive emotion experienced in response to an undervaluing of one’s welfare by another individual (Ortony, Clore, & Collins, 1988; Sell, Tooby, & Cosmides, 2009). Its function is to motivate the angry party to approach the angering parties and incentivize them to recalibrate their valuation (Carver & Harmon-Jones, 2009; van Honk & Schutter, 2005; Fischer & Roseman, 2007). These incentives— withholding cooperation, dissolving the relationship, or inflicting harm— must be honestly communicated to be effective. As such, expressions of anger communicate confrontational and unyielding states (Morris & Keltner, 2000; Hinde, 1985; Keltner & Haidt, 2001), and the intention to approach with aggressive intent (Yik & Russell, 1999; Adams, Ambady, Macrae, & Kleck, 2006).

Anger and threats often occur in bargaining situations when two individuals have conflicting preferences about the division of a resource, such as the sale price for a trade, but they also have a common interest in coming to an agreement rather than walking away from a deal. In this situation, individuals can use the threat of rejection to demand a particular share of the pie.

We use a modified ultimatum game to investigate the effects of angry expressions on the credibility of threats in bargaining. In the standard ultimatum game, a Proposer decides how much money to offer a Responder, the Responder decides whether to accept or reject the offer, and if the offer is rejected then both players receive nothing (Guth, Schmittberger, & Schwarze, 1982). Though a rational, self-interested Responder should accept any offer greater than zero, Responders are commonly “spiteful,” rejecting offers they perceive as unfairly low. This has been interpreted as a manifestation of anger (Straub & Murnighan, 1995), which is supported by the finding that rejections are accompanied by activity in brain areas associated with negative emotion (Sanfey, 2003). In turn, it is commonly thought that Proposers are more generous than rationality predicts because they anticipate and avoid provoking the Responder’s anger. This idea implies that when communication is possible Responders can gain by credibly communicating their anger, yet the effectiveness of these signals has not been demonstrated. The mere existence of an expression associated with anger is not sufficient, since these expressions may be evolutionary vestiges, spillovers of arousal, or cheap talk. To test the credibility of angry expressions, we altered the ultimatum game such that the Proposer viewed a threat issued by a confederate Responder (Experiment 1) or typical Responder (Experiment 2) before making their offer, and this threat was paired with either an angry or a neutral facial expression.

Further, to examine whether the expression specifically changes the threat’s credibility (rather than changing mood or some other generic effect), we also manipulated the inherent credibility of the threat by varying the amount demanded by the Responder. In many bargaining situations, negotiators converge on a 50-50 split, either because it is perceived as inherently fair or it is a salient “focal point” (Schelling, 1960). Accordingly, research in ultimatum game experiments shows that responders commonly accept offers of 50% and higher whereas lower offers are increasingly likely to be rejected as they deviate downward from 50% (Camerer, 2003). For example, List and Cherry (2000) reported (for $20 stakes) approximately 20% rejections for offers of 40-49%, 30% rejections for offers of 30-39%, and 70% rejections for offers less than 25%. A threat by a Responder to reject a lowball offer and accept an
even split is thus inherently credible and needn’t be reinforced by an honest signal. In contrast, a Responder who demands 70% of the pot on pain of spitefully rejecting lower offers is not making an inherently credible threat, because they can expect the Proposer to see it as unfair, and because carrying out the threat is against the Responder’s own interests. To prove the threat is not a bluff, the Responder needs an additional guarantor. According to the credible signal hypothesis, that guarantor is anger, signaled by an angry expression, and it will be effective in inducing the Proposer to make a more generous offer. Specifically, proposers will make a higher offer when an inherently non-credible threat (a demand for 70% of the pot) is accompanied by an angry expression than when it is accompanied by a neutral expression, but will be unaffected by the expressions when they accompany an inherently credible threat (50%). The alternative hypotheses (vestige, spillover, and cheap talk) predict that anger expressions will not affect Proposer offers at all.

**Experiment 1**

**Method**

This experiment examined Proposer offers in an Ultimatum Game with Responder threats. Before making their offers, Proposers were told they would be paired with a specific videotaped Responder. Threats consisted of a brief video clip of either a neutral or angry facial expression paired with a written message demanding either 50% or 70% of the pie. This resulted in a 2 (expression) by 2 (demand) design.

**Participants and Procedure.** Eight-hundred and seventy participants were randomly assigned as Proposers or Responders on Amazon’s Mechanical Turk (Mturk), an online crowdsourcing website in which individuals sign up to complete tasks. It has been used in previous research in psychology and experimental economics and has a large and diverse subject pool (Buhrmeister, Kwang, & Gosling, 2011; Horton, Rand, & Zeckhuaser, 2011). The mean age (571 male, 299 female) was 30.21 (SD = 9.34). Participants were 79% Caucasian, 7% African-American, 11% Asian-American, and 3% other race.

Participants were given a consent form and a description of the procedure. Participants in the Proposer role read that they received a $1.00 endowment and could decide how much to offer to a Responder. This endowment is similar in size to other online studies using economic games and yields results comparable to those in laboratory settings with greater stakes (Amir, Rand, & Gal, 2012). The Responder would then decide whether to accept or reject the offer, and if it was rejected then both players receive $0. The actual monetary payoffs were jointly determined by the participant and a matched partner. In addition to earning $0.40 for completing the study, participants could earn up to an additional $1.00 depending on the decisions that they and their partner made. These additional earnings were paid using the “bonus” feature on Mturk. Participants answered three comprehension questions (correct responses were required for participation) and completed a single trial of the task (~5 minutes).

Prior to making their offer, Proposers viewed the threat of a confederate Responder who was said to have “provided messages and been videotaped.” The threat consisted of a 6 sec video clip and a written demand. We used video clips because they provide more information to perceivers than static images (Ambadar, Schooler, & Cohn, 2005). Each clip depicted the same female actor who had been instructed to create specific facial actions and expressions. The actor displayed either a neutral (no facial actions) or an angry expression (frowning eyebrows, intensely staring eyes, and a closed mouth with lowered corners) (Ekman & Friesen, 1975). At the conclusion of the clip, the video screen turned black. Each clip was recorded at 30 frames per second. This produced a set of 180 sequential 640 x 480 full-color images. This length and frame rate allowed for presentation of facial expressions similar in length
to the average 4-6 seconds reported for spontaneous expressions (Frank, Ekman, & Friesen, 1993; Schmidt, Ambadar, Cohn, & Reed, 2006).

The clips were paired with written demands for either an equal split of 50% each (the inherently credible threat) or an unequal split of 70% for the Responder and 30% for the Proposer. The demands were written as: “If you offer me less than 50/70 cents, I will reject your offer.” We chose a demand of 70% to represent a threat of borderline credibility. This was based on an initial pilot study on threats composed of neutral expressions paired with demands of 60% through 90% in 10% increments, which found that 70% demands elicited the most compliance.

Proposers specified their offers immediately following the presentation of threats. Proposers answered, “How much money do you choose to transfer to the Receiver (0-100 cents)?” Then they were asked to rate the Responder’s happiness and anger on 7-point scales with 1 indicating “not at all” and 7 indicating “extremely”.

Proposers were told that their Responder specified a cut-off amount representing the minimum amount they would accept without forcing both of them to forfeit the entire pot. In actuality, participants in the Responder role made their decisions for each possible Proposer offer using the “strategy method” (Selten, 1967), stating how they would react to each amount that a Proposer might offer. After data collection (i.e., not in real time), Proposers and Responders were randomly matched, the Responder's contingent choices were carried out for the specific Proposer offer, and the payouts calculated accordingly.

**Results and Discussion**

As a manipulation check, we first examined participants’ ratings of the Responders’ emotions. With both demand amounts, Proposers rated typical Responders in the anger video clip as more angry ($F(1, 431) = 17.87, p < .001$) and less happy ($F(1, 431) = 21.01, p < .001$) than those in the neutral video clip. The ratings of anger were not significantly affected by the size of the demand ($F(1, 431) = 4.43, p = .036$), nor did demand size interact with expression ($F(1, 431) = 0.28, p = .601$). Ratings of happiness significantly decreased with the size of demand, $F(1, 431) = 7.95, p = .005$, though demand size did not interact with expression, $F(1, 431) = 0.86, p = .769$. Turning to Responders’ decisions, the average minimum acceptable amount specified was 31.15 ($SD = 19.32$). Last, Proposer offers did not differ by sex or ethnicity (all $p$'s > .05) so we aggregate across these categories in subsequent analyses.

The data of interest are shown in Figure 1. When Proposers read demands of 50%, they gave no higher offers when they viewed angry expressions ($M = 49.98, SD = 10.95$) than when they viewed neutral expressions ($M = 50.55, SD = 11.17$). In contrast, when they read demands of 70%, they gave significantly higher offers when they viewed angry expressions ($M = 60.92, SD = 11.23$) than when they viewed neutral expressions ($M = 54.77, SD = 20.73$) $t(215) = -2.71, p = .007$. A 2 X 2 ANOVA revealed significant main effects for facial expression, $F(1, 431) = 4.19, p = .041$ and demand, $F(1, 431) = 30.96, p < .001$, and crucially, a significant interaction between facial expression and threat, $F(1, 431) = 6.08, p = .014$. 
In sum, the data support the predictions of the credible signal hypothesis. When faced with a dubious demand of 70% of the pot, Proposers who viewed an angry expression offered more money than those who viewed a neutral expression. When faced with a more credible demand of 50%, an extraneous guarantor was otiose, and the Proposers were unaffected by the expression.

**Experiment 2**

**Method**

In Experiment 1, we used deception by telling participants that confederate angry and neutral videos showed their actual partners. It is possible that angry faces caused inferences about the specific partner (e.g., a malevolent disposition) rather than operating only as signals of credibility. In Experiment 2, we test the same effect without deceiving participants about their partner. Instead, the person in the video is described as a typical Responder rather than the Proposer’s specific partner. We test whether an angry expression increases the perceived credibility of threats even when it does not provide information about a specific partner.

**Participants.** Five hundred eighty-eight participants on Mechanical Turk were randomly assigned as Proposers or Responders. The mean age was 30.80 (SD = 10.08); 345 were male. Participants were 79% Caucasian, 7% African-American, 12% Asian-American, and 2% other race.

**Procedure.** The procedure, threats, and emotion ratings were identical to Experiment 1 with one exception: rather than viewing a confederate responder, participants viewed Responder threats that were described as “typical of a Responder in this scenario” and they were told they would be paired
with a randomly selected Responder. Following data collection, Proposers and Responders were randomly matched, and the Responder’s contingent choice was carried out for the specific Proposer offer made in their interaction, and the payouts calculated accordingly.

Results and Discussion

As in Experiment 1, a manipulation check confirmed that Proposers rated Responders in the anger video clip as more angry ($F(1, 290) = 31.39, p < .001$) and less happy ($F(1, 290) = 18.41, p < .001$) than Responders in the neutral clip. The ratings of anger were not significantly affected by the size of demand ($F(1, 290) = 3.32, p = .068$), nor did demand size interact with expression ($F(1, 290) < 0.001, p = .995$). The same pattern held for ratings of happiness: no significant main effect of demand, $F(1, 290) = 1.04, p = .308$, and no significant interaction, $F(1, 290) = 1.03, p < .310$. Turning to Responders, the average minimum acceptable amount specified was 33.53 ($SD = 17.45$). Last, Proposer offers did not differ by sex or ethnicity (all $p$’s > .05), so we aggregate across these categories in subsequent analyses.

The results of interest are depicted in Figure 2. When Proposers read demands of 50%, they gave virtually identical offers when viewing angry ($M = 46.07, SD = 16.11$) and neutral expressions ($M = 46.12, SD = 11.64$). In contrast, when they read demands of 70%, they gave significantly higher offers when they viewed angry expressions ($M = 54.04, SD = 15.41$) than when they viewed neutral expressions ($M = 46.13, SD = 17.24$), $t(145) = 2.91, p = .004$. The pattern is confirmed in a 2 X 2 ANOVA which revealed significant main effects for facial expression $F(1, 290) = 4.86, p = .028$ and demand $F(1, 290) = 5.02, p = .026$, and a significant expression by threat interaction $F(1, 290) = 4.99, p = .026$.

![Figure 2](image-url)

*Figure 2.* Mean (SE) Proposer offers in response to generic Responder threats accompanied by neutral or angry expressions and demanding 50% or 70% (Experiment 2)
As with Experiment 1, the data confirm the predictions of the credible signal hypothesis: angry expressions increase offers for a demand of 70%, which is neither fair nor plausible and hence not inherently credible, but not for a 50-50 split, which is a natural bargaining equilibrium and hence a credible threat with no need for an external guarantor. Importantly, Experiment 2 shows that this effect occurs even when the person displaying the angry face is described as a typical Responder rather than the Proposer’s specific partner, and thus reflects a strategic response to an angry expression rather than information about a specific partner. At the same time, Proposer offers in the key condition (angry expression, 70 cent demand) were greater in Experiment 1 than Experiment 2, \( t(174) = 3.42, p > .001 \). This indicates, as expected, that angry threats were most effective when issued by a specific partner, even if the connection to a specific partner was not required.

**General Discussion**

Results of both experiments support the hypothesis that angry facial expressions function as credible threats in bargaining situations: angry expressions resulted in increased offers when paired with a less-than-reasonable and less-than-plausible demand, but not when paired with a fair, plausible, and hence credible demand. The effects were strong enough to induce proposers to make generous offers (ones that favored their partners over themselves, which is highly unusual in Ultimatum Games), and they occurred whether the perceivers viewed expressions by their specific Responder in the bargaining situation or by a generic Responder. These findings suggest that angry expressions are not cheap talk but are perceived as honest signals of a person’s willingness to carry out a mutually harmful threat. As such, the expression and perception of anger can function to avoid spiteful behavior in bargaining situations, benefiting both senders and receivers.

The data thus support the adaptive function of emotions proposed by Frank (1988) and Hirshleifer (1987) in which the expressive components of emotion serve as credible signals communicating subjective commitment. In this case, angry facial expressions serve as an honest signal that one is motivated to act against rational self-interest and reject low Proposer offers. But the logic may apply to positive emotions as well: other experiments suggest that smiles serve as credible signals of cooperative intent (Reed, Zeglen, & Schmidt, 2012), and may guarantee promises against the suspicion of rational defection and double-crossing in the same way that angry expressions guarantee threats against the suspicion of bluffing.

One surprising aspect of the Proposers’ responses is that they occurred even though the facial expressions were created by using a directed facial action task rather than being spontaneously emitted. In the actual context of the experiment, the expressions were in fact not honest but literally faked, yet the participants treated them as if they were honest. Future research can examine this issue, as well the sending of commitment signals, by observing Responder threats in face-to-face interactions allowing spontaneous expressions of emotion.

The fact that bargaining offers are mediated by internal emotional states speaks not just to theories of the adaptive function of the emotions and emotional expressions but also to the study of bargaining. Experienced negotiators know that negotiations are affected not just by rational self-interest but by the participants’ emotional reactions, including perceptions about how much each side values the other side’s welfare. This is why, for instance, prospective homebuyers are advised not to start with a lowball offer because it will “insult” the homeowner, and salespeople feign (and sometimes honestly develop) warmth and friendship to their customers to increase the likelihood of profitable sales. These responses appear to be irrational quirks, and are exploited by seasoned negotiators, but they may have
a rationale in the less structured negotiations that make up informal social cooperation. The study of the role of emotions in negotiation and bargaining thus offers the prospect of illuminating both phenomena.
References


Author Contributions

L. I. Reed, P. DeScioli, and S.A. Pinker developed the study concept and contributed to the study design. Data collection was performed by L. I. Reed and P. DeScioli. L. I. Reed and P. DeScioli performed data analysis and interpretation. L. I. Reed, P. DeScioli, and S. A. Pinker wrote the manuscript. All authors approved the final version of the manuscript for submission.