Original Article

Watch out! How a fearful face adds credibility to warnings of danger

CrossMark

Lawrence Ian Reed [a,*], Peter DeScioli [b]

[a] *Department of Psychiatry, McLean Hospital, Harvard Medical School, United States*
[b] *Department of Political Science, Stony Brook University, United States*

A R T I C L E   I N F O

A B S T R A C T

People display facial expressions of fear to communicate danger to others and sometimes to exaggerate danger to manipulate an audience. Here we test whether fear expressions add credibility to a speaker's warnings of danger. Participants played an incentivized lie detection game in which they guess whether a confederate partner is lying or telling the truth. Participants viewed a video of their partner's message, after reading that there was a good chance (75%) their partner was instructed to lie. We manipulated across conditions whether the partner stated the message with a neutral or fearful expression. Experiment 1 finds that participants were more likely to believe the speaker's warning of danger when it was conveyed with a fear expression compared to a neutral expression. Experiment 2 finds that when a speaker instead claimed that a danger was absent, a fearful expression no longer added credibility to their message. These findings provide evidence that fear expressions add credibility to statements of danger, specifically, rather than any claim.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

A fearful face is an arresting signal of danger (Ekman, 1992; Vuilleumier, 2002). People direct their fear expressions toward other people to alert them of threats. Politicians, for example, often warn citizens about grave dangers, whether genuinely or manipulatively (Lupia & Menning, 2009), and they use fear expressions to augment their message. Similarly, fear expressions are used by public health officials to warn against threats like spreading viruses, by religious leaders to exhort followers to avoid supernatural perils, and by law enforcement to stress the menace of rising crime.

However intuitive it might be that people use fear expressions to communicate danger, there is little evidence that it actually works. And there are even good reasons to think it might not, at least in some cases. Namely, a fearful speaker could be exaggerating the danger or outright lying to manipulate the audience. This problem of credibility is posed by the classic folk tale *The Boy Who Cried Wolf* in which a boy who feigned danger is no longer believed. The folk tale *Chicken Little* also underscores the unreliability of fear when a chick struck by an acorn panics and proclaims, "The sky is falling!" In politics, both commentators and researchers have argued that citizens should pay less heed to doomsday claims by politicians and the media (Furedi, 2005; Glassner, 1999). If fear is used to mislead and manipulate, then fearful expressions might also elicit skepticism in addition to concern.

Theories of communication in game theory and evolutionary biology elaborate further on this problem of credibility. When a signal could benefit the sender at the recipient's expense, the recipient should eventually come to ignore the signal, or at least to discount its veracity (Dezecache, Mercier, & Scott-Phillips, 2013; McCullough & Reed, 2016). From this perspective, a fearful expression is cheap talk, a signal that could be faked to trick the recipient. Hence, people might ignore mere facial expressions and instead focus their evaluations on other information that is more difficult to fake such as concrete evidence of danger.

At the same time, however, there are countervailing reasons why people might still be swayed by a fearful face. Some evolutionary researchers have argued that emotional expressions were sculpted by the evolutionary process specifically to overcome the problem of credibility (Frank, 1988; Reed, DeScioli, & Pinker, 2014). The argument is that some emotional expressions are actually designed to be involuntary and difficult to consciously fake so that they convey more credible signals.

This model does confront a difficult theoretical hurdle: It seems that evolution would favor a mutant design that could fake the same expressions. However, there are some possible resolutions. For instance, it might be that individuals who are too good at faking emotions are eventually revealed as cold and calculating liars after repeated interactions in tight-knit social groups (McCullough & Reed, 2016; Searcy & Nowicki, 2005). If so, they could eventually suffer extreme social costs like exclusion, punishment, and ostracism, which would preserve the advantages of genuine facial expressions.

A second possibility is that fearful expressions remain credible because they are downstream effects of an involuntary fear response. It is arguably adaptive for the emotion of fear to remain resistant to

---

conscious control so that individuals cannot easily override their fear of grave dangers such as heights, crocodiles, or menacing strangers. Fearful expressions might be contained within the same cognitive architecture of fear which remains adaptively insulated from voluntary control. Although it is theoretically possible that natural selection could shape a counterfeit fear expression, this possibility could be blocked by developmental constraints. By analogy, it is theoretically possible to reroute the optic nerve to remove the blind spot in the vertebrate eye, but this has not occurred due to developmental constraints, despite its potential advantages (Williams, 1998). Similarly, there might be no simple way for natural selection to incrementally reroute fearful expressions to be under conscious control.

Amid these theoretical debates, researchers have also used empirical studies to investigate whether people do in fact believe claims backed up by emotional expressions. The credibility hypothesis (Reed & DeScioli, 2015; Reed et al., 2014) states that facial expressions function, in part, to add credibility to accompanying messages. Previous research has tested this theory for several different emotions and types of messages. For instance, several studies have found that smiles increase the credibility of promises to cooperate (Brown & Moore, 2002; Brown, Palameta, & Moore, 2003; Reed, Zeglen, & Schmidt, 2012). Another study found that angry expressions increase the credibility of threats in ultimatum bargaining (Reed et al., 2014). And a third study found that sad expressions increase the credibility of statements of loss (Reed & DeScioli, 2015).

Fear expressions might also serve a credibility function. Specifically, we propose that fear expressions bolster claims of danger. The emotion of fear is elicited by threats to oneself and others (Ekman, 1992) and is composed of largely automatic, unconscious, and involuntary processes (Kandel, Schwartz, & Jessell, 1995; LeDoux, 1996; Ohman & Mineka, 2001; Witte, 1992). The basic evolved fear mechanisms function to prepare the individual to escape the threat (Epstein, 1972). To this end, fear activates a suite of physiological responses including those affecting adrenalin, heart rate, and musculature (Ekman, 1992; Marks, 1987; Tooby & Cosmides, 2016). Many of these basic processes are shared widely across other mammalian species (Waller & Michelatta, 2013).

In some animal species, fear also controls a species-typical facial expression. In humans, this is characterized by raised inner and outer eyebrows, widened eyes, an outward pull of the lip corners, and dropped jaw. Darwin proposed that the fear expression directly helps an individual evade a threat (Darwin, Ekman, & Prodger, 1998), which is supported by some modern research. Raising the eyebrows and eyelids increases the size of the visual field and increases saccadic velocity. The outward pull of the lip corners and dropped jaw increase inspiratory capacity (Susskind et al., 2008). Together, these physiological changes prepare the individual to respond adaptively to threats (Susskind & Anderson, 2008).

These physiological changes may also serve a communicative function: to alert others of an imminent danger. Compared to an alarm call or scream, fearful expressions can communicate dangers quietly, acting as a silent warning that safely communicates danger without putting the signaler at risk. A signaling function is consistent with research showing heightened amygdala activity in response to viewing fearful faces (Adolphs, Russell, & Tranel, 1999; Morris et al., 1996; Whalen et al., 2001), which suggests that receivers of fear expressions automatically activate their own fear response. It is also consistent with the finding that children and adults associate someone else's fear expression with dangers like consuming poisonous foods or cleaning products, and they expect a person who is presented with danger to show an expression of fear, more so than other emotions (Pooley, Hon, & Fiddick, 2010). There is also evidence that viewing fearful faces affects behavior in adaptive ways. In a classic study, Sorce, Emde, Campos, and Klinnert (1985) found that infants frequently referenced their mother's facial expression before crossing a visual cliff of uncertain height. They found that few infants crossed when mothers posed fearful facial expressions, suggesting that the infants used these expressions to help assess the danger ().

Previous research has not, to our knowledge, specifically tested whether a fear expression affects the credibility of a signaler's claim of danger. In situations where the signaler and receiver share interests, as in the visual cliff study, credibility is not a problem. Receivers have no reason to distrust the message because signalers have no reason to be dishonest (McCullough & Reed, 2016; Searcy & Nowicki, 2005). However, when signaler and receiver could have conflicting interests, credibility becomes a potential problem. If the signaler displays a fear expression, the receiver cannot be sure if the signaler is telling the truth or trying to deceive them. It is currently unknown whether fear lends credibility even when there could be conflicting interests. Yet, this exact situation frequently applies, such as when an auto mechanic, lawyer, doctor, or politician offers costly measures to protect against a danger they claim is imminent.

The credible danger hypothesis predicts that fear expressions add credibility to messages even in situations with potentially conflicting interests. This does not of course mean that receivers will always believe fearful claims, but rather that, all else equal, a fearful claim is more believable than the same message conveyed without fear, even if both versions are viewed as relatively reasonable or dubious. The rationale for this idea is that fearful expressions evolved to be difficult to consciously fake, hence preserving their signal value. In response, receivers can use genuine fear expressions, among other cues, to assess the likelihood of danger and adjust their behavior accordingly. This hypothesis predicts that receivers will be more likely to believe claims of danger when they are paired with a fearful expression in comparison to a neutral expression. An alternative cheap talk hypothesis predicts that messages will be seen as equally credible whether paired with fearful or neutral expressions.

2. Experiment 1

2.1. Method

2.1.1. Participants

We recruited 218 participants (140 male, 78 female) using Amazon's MTurk, an online crowd-sourcing website where individuals sign up to complete tasks for payment. It has been used in previous research in psychology and experimental economics and has a large and diverse subject pool (Horton, Rand, & Zeckhauser, 2011; Buhrmeister, Kwant, & Gosling, 2011). Participants' mean age was 37.0 ($SD = 11.9$); their racial distribution was: 82.6% Caucasian, 8.7% African American, 6.9% Asian American and 2% other.

2.1.2. Lie detection task

We designed an incentivized lie detection task to observe participants' judgments about the credibility of a speaker's claim of danger. Participants viewed a pre-recorded video of a speaker stating a message about danger. Participants read that the speaker's message could be true or false and their task was to judge whether the speaker was telling the truth. If they guessed correctly, they would earn additional bonus money (50 cents). In reality, the message was recorded from a confederate actress and was neither true nor false, and participants were always paid the bonus.

We created a potential conflict of interest between the participant and the speaker by telling participants that there is a good chance the speaker is lying. Participants read that some speakers were instructed by the experimenter to lie, and specifically that there was a 75% chance that their partner was instructed to lie. We chose a 75% chance of conflicting interests so that participants would start with initial skepticism toward the speaker, allowing us to test whether a fear expression helps overcome this skepticism.

We described the task to participants using a fictional backstory to provide a concrete context for the speaker's message that included an element of danger and a possible motive for deception. Participants