J Risk Uncertain
https://doi.org/10.1007/s11166-018-9276-4

Corporate apology for environmental damage

Ben Gilbert[1] · **Alexander James**[2] · **Jason F. Shogren**[3]

Abstract Apologies are a powerful way to restore trust and reduce punishment costs in bilateral settings. But what do we know about public apologies for large scale man-made disasters? Herein we report on results from an experiment with apologies in a multilateral setting: a firm-caused environmental disaster. Subjects read about an oil spill scenario, and learned whether the oil firm made a full apology, a partial apology, or no apology, and whether the firm had a good, bad, or no environmental reputation. A partial apology is one that fails to accept full material responsibility for damages, such as by shifting the blame to another party. We find that full apologies and better reputation reduce the demand for punishment. However, full apologies and reputation are substitutes, with reputation being significantly more important. Additionally, apologies do not reduce the demand for compensation and may increase it if the firm is clearly a bad actor, or if admission of guilt is the only information subjects have. Our results help explain corporate social responsibility investments and greenwashing, and why many public apologies over an environmental disaster are only partial apologies.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11166-018-9276-4) contains supplementary material, which is available to authorized users.

✉ Alexander James
alex.james@uaa.alaska.edu

[1] Division of Economics and Business, Colorado School of Mines, Golden, CO 80401, USA

[2] Department of Economics and Public Policy, University of Alaska Anchorage, 3211 Providence Drive Anchorage, AK 99508, USA

[3] Department of Economics, University of Wyoming, Laramie, WY 82071, USA

Published online: 27 March 2018

Keywords Corporate social responsibility · Greenwashing · Apologies · Environmental disaster · Cheap talk

JEL Classifications Q54 · L14

1 Introduction

> "The responsibility for safety on the drilling rig is Transocean. It is their rig, their equipment, their people, their systems, their safety processes."
>
> *—Tony Hayward, BP CEO*
>
> *CNN interview, April 28, 2010*
>
> "The explosion and fire aboard the Deepwater Horizon and the resulting oil spill in the Gulf of Mexico never should have happened — and I am deeply sorry that they did. None of us yet knows why it happened. But whatever the cause, we at BP will do what we can to make certain that an incident like this does not happen again."
>
> *—Tony Hayward, BP CEO*
>
> *U.S. House Testimony, June 17, 2010*

CEOs often make public apologies when their firm's products cause harm. These apologies are a powerful way to restore trust and reduce punishment costs in bilateral settings such as medical malpractice contracts (see McMichael et al. 2016), food safety scares, product tampering incidents, automotive scandals (e.g., VW emissions, Ford and Firestone tires, GM ignition switches, Toyota accelerator/braking risks), and major security/data breaches (e.g., Target, Equifax, Yahoo mail). Highly visible apologies from CEOs such as Apple's Steve Jobs and Facebook's Mark Zuckerberg can take many forms, and they play an important role in dealing with adverse risk outcomes (see for example Hearit 2006). An apology can save a company millions of dollars in punishment and gain more goodwill with the public.

But what do we know about consumers' desire for punishment and compensation given public apologies for large-scale multilateral, man-made disasters? Herein we report on results from a novel experimental survey focusing on apologies within a multilateral setting: a firm-caused environmental disaster. For example, the statements in the epigraph by former BP CEO Tony Hayward following the Deepwater Horizon oil spill differ in some obvious and some subtle ways. The obvious difference is that in the first statement, made soon after the April 20, 2010 explosion causing the spill, Hayward shifts the blame to Transocean, the owner of the drilling rig that exploded. In the second statement, Hayward apologizes for the spill and pledges to avoid future harm. More subtle is that the second statement is only a partial apology; Hayward does not accept full responsibility for the damage on behalf of his organization.

A similar and now somewhat infamous statement was made by Exxon Chairman Lawrence Rawl ten days after the 1989 Exxon Valdez spill in Prince William Sound:

"I want to tell you how sorry I am that this accident took place."

—Lawrence Rawl, Exxon Chairman & CEO

Newspaper ads, April 3, 1989

Could such statements have affected the demand for punishment and compensation from these firms? Rawl's statement was seen as stoking public outrage over the spill. The statement does not accept responsibility, was made with significant delay, and was made in a passive medium (newspaper advertisements) in which its sincerity could not be evaluated (O'Hara O'Connor 2011); news outlets reported that 40,000 customers subsequently cut up their Exxon credit cards in order to boycott the company (Behar 1990).

Conventional wisdom suggests that apologies help resolve conflict and restore trust at reduced cost. This view is supported by a broad literature in economics, law, management, communications, and psychology discussed in more detail below. But this standard view is based on bilateral interactions, which provide too narrow a perspective for large scale environmental disasters. Environmental damages are inherently multilateral; they involve the loss of a public good, with multiple firms or agents potentially at fault, and incidence of costs borne by a large group of victims. This multilateral setting may create strategic reasons for firms to make only "partial" apologies — such as shifting the blame to a third party — at the risk of being less effective.

This paper reports results from an experiment that tests how corporate apologies for environmental damage affect the demand for punishment and compensation. Specifically, we ask whether the apology's content, and the firm's reputation, affect the outcome. We study two types of punishment outcomes: (1) "personal responses", which include boycotting, opposing local development, and signing a petition to urge criminal prosecution of a firm, and (2) fines in excess of victim compensation. We also study the demand for compensation when the subject is part of a large group of victims, and compensation is determined through a hypothetical legal settlement. In a 3×3 design, we presented subjects with an oil spill scenario and randomly assigned them to one of three apology treatments (No Apology, Full Apology, Shift the Blame), and one of three firm reputation treatments (No Reputation Information, Good Reputation, Bad Reputation). We then asked subjects about their likely personal responses to the spill, their preferred fine, and their Willingness to Accept compensation (WTA) as part of a class action lawsuit for lost passive use value of the environmental good. Subjects were not told the scenario and firms were fictional until the survey was complete.

We find that full apologies and better reputation reduce the intensity of personal responses. Apologies and reputation seem to be substitutes in affecting personal responses, however, with reputation being significantly more important. The relationship becomes more complicated for the cases of the fine and WTA. For the fine, if the firm has a good reputation, people ask for a smaller fine — *especially* if the firm shifts the blame. For WTA, if the firm has a good reputation, then giving a full apology has no additional impact on WTA — the impact is the same if the firm remained silent. But we find weak evidence that a person's WTA *increases* if (a) the firm accepts blame, irrespective of the firm's reputation, or (b) the firm has a bad reputation and shifts

the blame. Subjects seem to either latch on to the admission of guilt, or they want to punish the clear bad actors. These results help explain why firms make investments in corporate social responsibility (CSR) that provide cover for socially irresponsible behavior, as in Kotchen and Moon (2012), or engage in "greenwashing", or the deliberate spread of misleading information about their environmental record (Cherry and Sneirson 2011; Delmas and Burbano 2011). The results also illustrate why many corporate apologies are often only "partial" apologies that shift the blame, downplay potential damages, or fail to fully accept responsibility, as was true with the apologies for both the Deepwater Horizon and the Exxon Valdez oil spills.

The rest of the paper is organized as follows. The next section discusses related literature on apologies from several disciplines. Section 3 describes the experiment and methods. Section 4 discusses the results, and Section 5 concludes.

2 Background

This paper contributes to a growing economics literature, and a large literature on apology and remorse outside of economics, that studies how apologies affect outcomes following economic harm. The bulk of this literature considers the effect of apologies in bilateral settings. Our paper is unique in that it tests the effect of an apology in a multilateral setting involving the loss of a public good, and it analyzes tradeoffs between apology content and objective reputation information. Our paper also distinguishes between punishment and compensation demanded as separate responses to harm. The present study is, to the best of our knowledge, the first to systematically investigate the impact of corporate apologies for environmental accidents on the punishments and compensation demanded by the public.

Most of the existing literature supports the general conclusion that apology messages can affect the punishments and compensation demanded by injured parties. Abeler et al. (2010), for example, find that customers who are disappointed with a purchase are more likely to forgive the seller if they receive an apology rather than a compensation payment. According to O'Hara O'Connor and Yarn (2002), 83% of lawyers believe an apology alone could settle disputes in many cases that are escalated to litigation in the absence of apology. Evidence from the economics and legal literature on civil litigation, medical malpractice, and even criminal cases indicates that an *effectively structured* apology can reduce liability, settlement amounts, malpractice claims and occasionally criminal sentences (Boothman et al. 2009; Cohen 1999; Ho and Liu 2011; Korobkin and Guthrie 1994; O'Hara O'Conner 2004, O'Hara O'Conner and Robbins 2009; O'Hara O'Conner and Yarn 2002; Pace et al. 2010; Robbennolt 2003). However, recent evidence by McMichael et al. (2016) suggests that apology laws related to medical malpractice may not be as effective as previously thought.

An "effective" apology recognizes shared norms between the transgressor and victim, expresses remorse at the violation of the shared norm, and promises to repair and avoid future offense (O'Hara O'Connor 2011). O'Hara O'Connor (2011) conducts a comprehensive review of the literature and concludes that "conciliatory efforts can be seen as a way to economize punishment costs in the face of defection". Her

review identifies four key components of an effective apology that persist across the literature in different disciplines:

1. Identify wrongful act and accept responsibility
2. Express remorse
3. Promise effort to avoid future damage
4. Offer resources to repair damage

Ineffective apologies are also marked by

- Delay
- Passive medium
- Unwillingness to accept burden

Economists have more narrowly considered apologies as signals in the tradition of Spence (1973), as in Ho (2011), Ho and Liu (2012), Martinez-Vaquero et al. (2015) and Ohtsubo and Watanabe (2009), for example. In this framework, apology messages need not be conciliatory — they need only be costly signals that a transgressor is a "good type" experiencing transitory bad luck.[1] These signals could include messages that are not apologies at all, including blaming others as a way to preserve one's reputation or the perception of one's type. By varying the apology messages in our experiment, we are able to study the effect of the apology's content, which sociologists, psychologists, and legal scholars argue is important for influencing outcomes. The content of these signals may differ in their expected cost to the sender, however, if accepting blame carries a greater liability risk than shifting blame — which is consistent with the signaling model.

The laboratory studies in economics and psychology tend to consist of bilateral pairs playing controlled bargaining, trust, and cooperation games over small sums of money; various apology messages with well-defined costs can be offered when one player deviates from cooperative or fair behavior. The general findings from these studies are that real information on an offender's innocent intentions and a credible apology can increase the chances that a victim will accept a given settlement, reduce the required punishment, and improve the return to trust and cooperative behavior (Bottom et al. 2002; De Cremer 2010; De Cremer et al. 2010, 2011; Desmet et al. 2011; Fischbacher and Utikal 2013; Gold and Weiner 2000; Ho 2012; Kim et al. 2004; Schniter et al. 2013).

In a trust game experiment, Fischbacher and Utikal (2013) investigate the receiver's inference problem by varying both the cost of the apology and the certainty of the offender's intentions. Intentions are controlled in their experiment by manipulating the difficulty of cooperation; defections that occur when cooperation is easy are interpreted as clearer evidence of bad intentions than defections that occur when cooperating is hard. Fischbacher and Utikal (2013) find that apologies reduce punishments when intentions are unclear, but that receivers punish apologizers when bad intentions are obvious (i.e., cooperating was easy but a defection occurred anyway). Our design is similar to Fischbacher and Utikal (2013) in that reputation may

[1] Mungan (2012) even designs a mechanism to explicitly price apologies in the legal system in order to give the signal a clear cost.

also signal intentions. Effectively, the receiver gets two signals of the sender's type, which may be substitutes: the apology and the reputation information.

Unlike the previous literature, however, with a corporate apology for environmental damage many agents may be to blame, punishment and compensation must satisfy the preferences of many heterogeneous victims, and apologies are made on behalf of a corporation rather than a person. The apology for the Exxon Valdez spill by Chairman & CEO Lawrence Rawl, for example, has been extensively analyzed in the communications literature[2] and has influenced the development of corporate "crisis management" as a subset of reputation risk management.[3] The effectiveness of a response to a public relations crisis may therefore depend on the firm's previous investments in reputation management. This can include misleading information as in the case of greenwashing as well as real investments in CSR. Greenwashing is implicitly incentivized in corporate law (Cherry and Sneirson 2011) as well as in market institutions (Delmas and Burbano 2011). Our results suggest greenwashing and CSR also provide ex ante crisis management insofar as they improve reputation. This is similar to the empirical results in Kotchen and Moon (2012), who find that firms that are more socially irresponsible also invest more in CSR.

3 Experiment and methods

The experiment was administered through an online survey in which all subjects were shown a common fictional oil spill scenario, and were then randomly assigned to one of three apology treatments combined with one of three reputation treatments in a 3×3 design. The treatments are described below. A sample of respondents that matches demographics of the 2010 U.S. Census was recruited from U.S. states in the Midwest by the online survey research firm Qualtrics.[4] Respondents were paid $5 for a completed survey. Respondents were not told the scenario was fictional until a debriefing message was offered at the end of the survey. The survey instrument is included in the corresponding electronic supplemental material.

The oil spill scenario described a spill on the coast of California near a small wildlife and marine reserve. We attempted to describe an event that was small enough

[2]See, e.g., Benoit (1997), Pauly and Hutchison (2005), Small (1991), Tyler (1997), Williams and Treadaway (1992).

[3]Effective reputation and crisis management is costly and valuable. Corporate crisis management specifically has been associated with long term stock price impacts (Knight and Pretty 1996). A more recent Deloitte survey of global executives found that reputation risk is rated as the most important strategic business risk, with growing importance as firms surveyed plan to invest more in reputation and crisis management, and allocate direct responsibility for reputation risk at the chief executive and board level (Deloitte 2014).

[4]We recruited subjects from: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. We purposefully chose to limit our subject pool to avoid surveying residents of the West Coast who would be more likely to know about any local oil spill. We acknowledge that this geographic targeting limits the external validity of our experiment. Future work that interacts our treatments with a wider array of subject characteristics (including, for example, a person's proximity to the impacted area) could be useful.

not to have been widely reported in the Midwest but damaging enough that Midwest residents might have some passive use value for the damages. We described an area with sensitive habitat and some endangered species and showed pictures of species that lived in the reserve. It is important to note that we were not interested in eliciting precise passive use values for the environmental amenities in this fictional reserve, but in measuring the difference in desired punishments for damage to the reserve, and behavior in a legal setting, across apology signals and reputation information.

The apology messages included a "No Apology'' message, a "Full Apology" message, and a "Shift the Blame" to a third-party message. The apology messages were designed using the components of effective and ineffective apologies identified in O'Hara O'Connor (2011), discussed in the previous section. The reputation treatments included "No Reputation Information", a "Good Reputation" message with objective information that the firm has previously been a good environmental steward, and a "Bad Reputation" message with objective information that the firm has multiple previous environmental violations.

Our "Full Apology" treatment contains all four components of O'Hara O'Connor's (2011) effective apology and none of the ineffective components, while our "Shift the Blame" treatment includes the word "apology" but leaves out key elements for an effective apology and includes some of the ineffective elements.

The "Full Apology" was stated as follows:

> "On behalf of our management team, I would like to convey our deep remorse over the damage this spill has caused to our environment and extend our sincerest apology. We are disappointed by this lapse in our safety protocol and we are adjusting our procedures to minimize the chances of, and impacts from, future spills. We would like to pledge whatever resources we can to assist in the cleanup and plan to open a fund to cover the damages."

The "Shift the Blame" message was stated as follows:

> "On behalf of our management team, I would like to extend our sincerest apology for this incident. We are investigating an engineering contractor whose negligence we believe is at fault for the spill. We have also sent a clean-up crew that will be sufficient to remove the oil, although damages will be minimal because the spill covers such a small area. Again, we apologize and hope that this matter can be resolved swiftly at the least cost to all parties."

Notice that the full apology includes the four elements of an effective apology outlined by O'Hara O'Connor (2011): the wrongful act is correctly identified, remorse is expressed, resources are offered for repair and a pledge is made to avoid future damage. The Shift the Blame message, on the other hand, blames an engineering contractor, attempts to downplay the damages, and emphasizes minimizing costs rather than repairing damages.

The "Good Reputation" treatment was stated as follows:

> "In the last 10 years, this company has had no other sizeable oil spills (one of the lowest rates in the industry), and they have won awards from multiple local

communities for good stewardship. After the recent spill, the firm sent a large force of clean-up workers and worked hand-in-hand with local volunteers and nonprofit groups."

The "Bad Reputation" treatment was stated as follows:

"In the last 10 years, this company has had more than 20 spills of at least 50 barrels of oil (one of the highest rates in the industry), and they have appeared on multiple watchdog groups' "worst of the worst" lists for their handling of environmental accidents. After the recent spill, the number of clean-up workers sent by the firm was not sufficient to remove the oil and clean up was handled primarily by local volunteers and nonprofit groups."

Notice that objective information in the Bad Reputation treatment conflicts with statements in the Shift the Blame signal, so respondents that received the combined treatment were told that the firm claimed to send a sufficient clean-up crew, but that the firm's resources were in fact not sufficient.

All subjects were told that the firm has already paid mandated clean-up costs and compensation for those directly affected by the spill, so that the effects we measure are for punishment and compensation above and beyond the firm's direct responsibilities. We evaluate the effect of apologies and firm reputation by estimating the impact of these treatments on the likelihood of punitive personal responses, the preferred federal fine size, and the willingness to accept a compensation settlement in a hypothetical class action lawsuit. We also gathered information about general demographics and environmental attitudes such as whether the subject had ever visited an ocean or a national park, their self-reported environmental preferences on a scale of 1 to 5, the presence of children in their household, their age, and their income. Table 1 presents the summary statistics for these variables. Income was measured in $25,000 increments on a scale of 1 to 8, so the table indicates the average income was a little more than $50,000.

3.1 Personal responses

Subjects were asked on a seven-point Likert scale how likely they would be to "boycott the firm's products", "oppose local development projects if this company is involved", or "sign a petition urging federal prosecution of this company, if asked".

We estimate the following model by ordered logit:

$$y_i = \beta_0 + \beta_T Treat_i + x_i'\gamma + u_i, \tag{1}$$

where y_i is the Likert scale response for subject i, $Treat_i$ is an indicator variable for the combination of apology and reputation messages shown to subject i, and x_i is the vector of control variables described above, including age, age squared, income bracket, environmental sentiment, and dummy variables for whether the subject has children, has ever visited an ocean, or ever visited a national park.

Table 1 Summary statistics

Variable	N	Mean	Std. Dev.	Min	Max
Baseline	750	0.109	0.312	0	1
Sorry	750	0.121	0.327	0	1
Blame	750	0.125	0.331	0	1
Good	750	0.113	0.317	0	1
Good, Sorry	750	0.100	0.300	0	1
Good, Blame	750	0.105	0.307	0	1
Bad	750	0.100	0.300	0	1
Bad, Sorry	750	0.123	0.328	0	1
Bad, Blame	750	0.103	0.304	0	1
Visit ocean	747	0.210	0.408	0	1
Visit park	747	0.763	0.425	0	1
Environmentalism	747	2.50	0.842	1	5
Age	750	49.8	14.1	19	84
Children Dummy	750	0.353	0.478	0	1
Income Category	746	2.61	1.48	0	8
Apology Matter (% Yes)	508	0.307	0.462	0	1
Smaller Fine	31	0.199			
Larger Fine	33	0.212			
No Difference	85	0.545			
Not Sure	7	0.449			
Smaller WTA	40	0.270			
Larger WTA	24	0.162			
No Difference	71	0.480			
Not Sure	13	0.878			
Information Matter (% Yes)	483	0.718	0.450	0	1
Smaller Fine	91	0.262			
Larger Fine	168	0.484			
No Difference	72	0.208			
Not Sure	16	0.461			
Smaller WTA	77	0.232			
Larger WTA	131	0.395			
No Difference	103	0.310			
Not Sure	21	0.633			

3.2 Preferred fine size

Subjects were asked how large of a fine the oil company should pay *in addition* to the cleanup costs. Dollar amounts were given in a sequence of dichotomous choices between \$0 and "more than \$15 million".

We estimate the following model by ordinary least squares with the "more than \$15 million" response hard coded as \$17.5 million, a tobit regression with the dependent variable censored at \$15 million, and ordered logit:[5]

$$f_i = \beta_0 + \beta_T Treat_i + x_i'\gamma + u_i, \tag{2}$$

where f_i is the fine preferred by subject i. We report results with and without the vector of control variables x_i.

3.3 Willingness to accept

Before beginning the survey, subjects were asked which of the two statements best describes them: "I do not value the conservation of marine species and habitats at all" versus "I have at least some value for the conservation of marine species and habitats in U.S. waters". Subjects agreeing with the first statement (5.3%) were coded as having a willingness to accept compensation of \$0. The remaining 94.7% of subjects who agreed with the second statement were asked to consider themselves part of a settlement negotiation for compensation for those indirectly affected by the spill. In a double-bounded dichotomous choice framework, subjects were asked if they would vote to accept the settlement if all eligible parties, including the subject, would be compensated \$100. Subjects who said "yes" were then asked if they would vote to accept \$50, whereas subjects who said "no" to \$100 were asked if they would vote to accept \$200.

It is important to note that this is not an incentive compatible elicitation mechanism for a subject's true, lost passive use value associated with the damaged marine reserve. Our results should therefore be interpreted as predicted behavior during a settlement negotiation. These estimates are nonetheless policy relevant, particularly for environmental damage lawsuits.

Following the standard approach in the literature to estimating WTA from a double-bounded dichotomous choice procedure,[6] if we let v_i, the compensation offer that subject i is just willing to accept, be logistically distributed, then the probabil-

[5]OLS results are robust to using different numbers of the top-coded response. Additionally, preferences for environmental fines may not be monotonic in the population, so we also estimated the effects on fine size using an unordered discrete choice conditional logit model. In the conditional logit model, the left hand side variable is equal to one for subject i's preferred fine and zero for every other fine option, and the right hand side variables include a continuous variable for fine size in addition to interactions between the continuous fine variable and the treatment dummies. Results from this model confirm those reported here and are available in the corresponding online appendix.

[6]See, e.g., Haab and McConnell (2002), Park et al. (1991).

ity of a "yes" vote at a particular compensation offer (or "bid") is related to the bid, treatment, and other covariates according to

$$ln\left[Pr(yes)/(1-Pr(yes))\right] = \beta_0 + \beta_b bid + \beta_T Treat_i + x_i'\gamma + u_i. \quad (3)$$

We estimate the parameter vector $\theta = (\beta, \gamma)$ using the log-likelihood function,

$$lnL(\theta) = \sum_{i=1}^{N} \left\{ d^{nn} ln Pr(v_i \geq 200) + d^{ny} ln\left[Pr(v_i \leq 200) - Pr(v_i \leq 100)\right] + d^{yn} ln\left[Pr(v_i \leq 100) - Pr(v_i \leq 50)\right] + d^{yy} ln Pr(v_i \leq 50) \right\}. \quad (4)$$

We then calculate the mean WTA and median WTA for each treatment T using the formulas below, where $\overline{x}$ is the vector of sample means of the control variables.

$$\begin{aligned} Mean\ WTA_T &= 1/\beta_b \left(ln(1 + exp(\beta_0 + \beta_T + \overline{x}'\gamma))\right) \\ Median\ WTA_T &= \frac{\beta_0 + \beta_T + \overline{x}'\gamma}{\beta_b} \end{aligned} \quad (5)$$

Confidence intervals for these estimates were simulated using the method of Krinsky and Robb (1986).[7]

4 Results

In this section we discuss the results of the experiment for personal responses, size of the preferred fine, and WTA. Having established how apologies affect punishment and compensation-seeking, in Appendix A we also compare these effects to how subjects *think* they respond to an apology — which can deviate from the actual treatment effects.[8]

4.1 Personal responses

Table 2 reports results from an ordered logit model for the effect of apology and reputation treatments on individual acts of retribution against the firm, measured by a

[7]We use Wilner's (2007) Stata package "wtpcikr" for simulating Krinsky and Robb confidence intervals, which is based on Krinsky and Robb (1986) as well as Loomis and Ekstrand (1998) and Poe et al. (2005). When confidence intervals are a nonlinear function of estimated standard errors, this method provides better estimates than the delta method. The method uses the estimated covariance matrix from the maximum likelihood model to simulate 5000 predicted mean and median WTA for different random draws of the error, and then chops off the smallest and largest 2.5% of the draws.

[8]In debriefing questions, subjects were strongly split on whether they thought the apology affected their preferred fine and WTA, based on whether the firm's reputation was good or bad. The subset of subjects who believe the apology was influential for them tended to believe it reduced their WTA and preferred fine. This is, however, not entirely consistent with how this subset actually responded to the treatments. Appendix A has a full discussion of these results.

Table 2 Treatment effects on likelihood of a personal response

	Boycott Products	Oppose local Development	Sign Petition for Prosecution
Sorry	0.21	0.34	0.38
	(0.262)	(0.252)	(0.281)
Blame	−0.16	0.017	−0.046
	(0.276)	(0.273)	(0.274)
Good	−0.49*	−0.52*	−0.30
	(0.286)	(0.275)	(0.294)
Good, Sorry	−0.89***	−0.68**	−0.73**
	(0.310)	(0.310)	(0.301)
Good, Blame	−0.49*	−0.40	−0.43
	(0.274)	(0.262)	(0.272)
Bad	0.68**	0.75***	1.06***
	(0.281)	(0.280)	(0.286)
Bad, Sorry	0.54**	0.57**	0.52*
	(0.275)	(0.275)	(0.283)
Bad, Blame	0.67**	0.67**	0.75***
	(0.296)	(0.279)	(0.281)
Visit ocean	0.20	0.26	0.32*
	(0.190)	(0.185)	(0.183)
Visit park	−0.17	−0.14	−0.25
	(0.155)	(0.154)	(0.158)
Enviro	0.66***	0.66***	0.69***
	(0.0937)	(0.0956)	(0.0958)
Age	0.061**	0.087***	0.052*
	(0.0295)	(0.0302)	(0.0314)
Age^2	−0.70**	−0.92***	−0.61*
	(0.299)	(0.312)	(0.324)
Kids	0.16	0.20	0.038
	(0.154)	(0.153)	(0.154)
Income	−0.054	−0.073	−0.054
	(0.0486)	(0.0524)	(0.0503)
N	741	741	741
Pseudo R^2	0.048	0.048	0.051

Note. Respondents stated the likelihood that they would engage in each action on a seven-point Likert scale. Regressions are estimated by ordered logit, but results are robust to ordered probit and OLS estimation and those results are available in the corresponding online appendix. Treatment dummies are relative to the baseline (No Apology, No Reputation) treatment. Robust standard errors are given in parentheses below regression coefficients, with statistical significance indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The Age^2 variable was divided by 1,000 to rescale its coefficient

Likert scale indicator for the likelihood of boycotting the firm's products (column 1), opposing local development by the firm (column 2), and signing a petition urging criminal prosecution of the firm (column 3).[9] Because the effects of treatment are similar across the three types of personal responses, we will refer to a "personal response" as one of these outcomes that occurs outside the formal (hypothetical) judicial system. Each of the treatment coefficients represents the average change in the log odds of moving to a higher Likert scale likelihood of a personal response, relative to the baseline group of No Apology or blame signal and No Reputational information. The "Sorry" and "Blame" coefficients are the effects of receiving an apology or blame signal, respectively, while not receiving any objective reputational information. The "Good" and "Bad" coefficients are the effects of receiving objective reputational information, while not receiving an apology or blame signal. The remaining treatment coefficients are the effects of receiving a combination of a particular apology and reputational information. For example, the "Good" coefficient of -0.49 in column 1 of Table 2 indicates that the odds that a subject would change from "Undecided" about boycotting to any higher category (Somewhat Likely, Likely, or Very Likely) *falls* by 49% if the subject receives the Good Reputation, No Apology treatment (relative to the No Apology, No Reputation baseline). The identical coefficients on "Good" and "Good, Blame" in column 1 indicate that blame shifting has no additional effect when the firm already has a good reputation. As another example, the "Bad, Blame" coefficient of 0.67 in column 2 indicates that the odds that a subject would change from "Unlikely" to oppose local development to any higher category (Somewhat Unlikely, Undecided, Somewhat Likely, Likely, or Very Likely) *increases* by 67% if the subject receives the Bad Reputation, Shift the Blame treatment (again relative to the No Apology, No Reputation baseline).

Across the personal responses, a good (bad) firm reputation decreases (increases) the inclination toward punitive personal responses. The coefficients on all treatments with good reputational information are negative and often statistically significant, whereas the coefficients on all treatments with bad reputational information are positive and always statistically significant. A full apology is a more effective way to augment a good reputation, and to alleviate a bad reputation, than is shifting the blame. For example, the combined effect of a good reputation and a full apology is always significant across all personal responses and is larger in magnitude than a good reputation alone, or a good reputation with a blame-shifting message. In addition, the effect of bad reputational information is smaller (although not statistically significantly) when combined with a full apology than when presented by itself or with blame shifting. In general, blame shifting has little discernible effect on the likelihood of a personal response.[10] As an important caveat, the differences in the coefficients obtained by adding an apology message to a reputational treatment are not statistically significant; for the majority of treatment combinations and personal

[9] In unreported results, we also estimated these effects by ordered probit and OLS with similar results which are available in the corresponding online appendix.

[10] The sole exception being the likelihood of signing a petition. For this outcome, shifting the blame reduces the likelihood of a personal response relative to the Bad Reputation, No Apology treatment.

response types, a statistically significant change relative to the baseline is driven by the reputation effect Table 3.[11]

Analyzing the additional controls reveals that subjects that think of themselves as environmentalists (either very strongly or somewhat so) are more likely to support personal responses. Age has a quadratic relationship in which the likelihood of a personal response peaks around age 45. The remaining controls mostly enter the model insignificantly.

We also find strong differences in response to the apologies and reputation information depending on whether subjects 1) are environmentalists and 2) have ever visited an ocean.[12] Tables 4, 5 and 6 report results on these subsamples for the three personal responses: boycotting, opposing, or petitioning for criminal prosecution of the firm. We find that non-environmentalists and subjects who had never visited an ocean were most likely to reward a full apology combined with a good reputation, and most likely to punish a firm with a bad reputation. On the other hand, environmentalists and subjects who had previously visited the ocean were most likely to punish a firm with a bad reputation who shifts the blame, but they respond comparatively less to the other treatments. One interpretation of these results is that environmentalists, and people with (perhaps) stronger preferences for marine environments, also have stronger prior beliefs and preferences and are less responsive to new information. Because of their potentially stronger preferences for marine environments, however, they are more likely to punish when the firm is clearly a bad actor.

4.2 Size of fine

Table 7 shows the effect of each treatment on the preferred fine, using several regression specifications. The first two columns report OLS regression results with and without control variables, with the preferred fine in millions of dollars as the dependent variable. The largest fine available in the survey was "More than $15 million", which we coded as $17.5 million for the results in the first two columns. Because we don't know the maximum preferred fine, the middle two columns report tobit regression results (with and without control variables) with the fine size censored at $15 million. For robustness, we also estimated these specifications using an ordered logit

[11] In a set of debriefing questions we asked subjects whether they believed the firm will change safety practices in the future. We estimate an additional logit model using this information. To the extent that subjects' beliefs about the firm's future practices reflect their trust in the firm, these results give a rough picture of whether the apologies and reputation information help restore or maintain trust. Table 3 shows the results which generally complement the findings for the personal responses. Bad reputational information decreases the likelihood that a subject reports believing the firm will change practices in the future, regardless of apology type. Apologies have no statistically significant impact, and therefore do not appear to help restore trust.

[12] Subjects were asked in a follow up survey, "Do you think of yourself as an environmentalist?" Subjects that responded "Yes, very strongly" or "Yes, somewhat" were coded as being environmentalists. We also separated the sample according to whether subjects had visited a national park. These results yielded largely insignificant and confounding results, perhaps due to the fact that so many people have visited parks, regardless of their preference for marine ecosystems. These results are available in the corresponding online appendix.

Table 3 Evaluating ex-post perception of offending firm

	Debriefing question: Will firm change safety practices?	
	Yes	No
Sorry	−0.34	−0.20
	(0.362)	(0.475)
Blame	−0.95**	−0.39
	(0.371)	(0.460)
Good	0.39	−0.99
	(0.386)	(0.652)
Good, Sorry	0.30	−0.77
	(0.381)	(0.620)
Good, Blame	−0.0019	−0.81
	(0.374)	(0.578)
Bad	−0.33	0.90*
	(0.426)	(0.481)
Bad, Sorry	−0.71*	0.70
	(0.394)	(0.444)
Bad, Blame	−0.56	0.85*
	(0.420)	(0.474)
Intercept	0.16	−0.53
	(1.046)	(1.343)
N	651	
Pseudo R^2	0.077	

Note. Estimates are from a multinomial logit model in which "Don't know" is the base response. Treatment dummies are relative to the baseline (No Apology, No Reputation) treatment. All specifications included control variables. Coefficients are suppressed for brevity but are available in the corresponding online appendix. Control variables were: ocean and national park visitation dummies, environmental sentiment, age, age squared, income, and children dummy. Fine and WTA are the subject's preferred fine size and WTA from the valuation task. Robust standard errors are given in parentheses below regression coefficients, with statistical significance indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

with discrete fine size options as the dependent variable. The results are robust across specifications, but the statistically significant "Tobit σ" parameter indicates that the tobit model is preferred to OLS.

The results suggest that shifting the blame, i.e., creating uncertainty about who was responsible for the loss, is a weakly better strategy for reducing fines than silence or full apologies in almost any reputation environment. The coefficients on the "Blame" treatments are lower for each of the three reputational environments than the respective reputation alone or the respective reputation combined with a full apology. However, the differences between these coefficients are not statistically significant. The exception to this pattern occurs in the Bad Reputation treatments, where

Table 4 Heterogeneous treatment effects: Boycott Products

	Enviro	Non Enviro	Visited Ocean	Never visited Ocean
Sorry	−0.12	0.49	−0.39	0.41
	(0.384)	(0.376)	(0.620)	(0.298)
Blame	0.013	−0.48	0.35	−0.23
	(0.389)	(0.392)	(0.650)	(0.312)
Good	−0.58	−0.54	0.81	−0.82***
	(0.423)	(0.383)	(0.762)	(0.317)
Good, Sorry	−0.77	−1.13***	0.0033	−1.17***
	(0.469)	(0.388)	(0.786)	(0.341)
Good, Blame	−0.61	−0.45	−0.37	−0.53*
	(0.394)	(0.394)	(0.676)	(0.307)
Bad	0.53	0.72*	0.40	0.84***
	(0.441)	(0.385)	(0.753)	(0.309)
Bad, Sorry	0.55	0.60	0.0072	0.75**
	(0.414)	(0.416)	(0.649)	(0.308)
Bad, Blame	0.71	0.52	1.19	0.59*
	(0.456)	(0.401)	(0.810)	(0.333)
Visit ocean	0.34	−0.22		
	(0.236)	(0.328)		
Visit park	0.18	−0.38*	0.59	−0.24
	(0.262)	(0.194)	(0.443)	(0.168)
Enviro	−0.46	−0.38**	−0.82***	−0.63***
	(0.284)	(0.182)	(0.214)	(0.107)
Age	0.10**	0.017	0.16*	0.037
	(0.0445)	(0.0383)	(0.0937)	(0.0306)
Age^2	−0.0010**	−0.00033	−0.0016	−0.00048
	(0.000450)	(0.000392)	(0.00103)	(0.000308)
Kids	0.22	0.22	0.20	0.19
	(0.221)	(0.223)	(0.387)	(0.175)
Income	−0.077	−0.040	−0.0038	−0.059
	(0.0670)	(0.0756)	(0.106)	(0.0561)
N	384	357	155	586
Pseudo R^2	0.033	0.038	0.062	0.055

Note. Respondents stated the likelihood that they would engage in each action on a seven-point Likert scale. Regressions are estimated by ordered logit, but results are robust to ordered probit and OLS estimation. Treatment dummies are relative to the baseline (No Apology, No Reputation) treatment. Robust standard errors are given in parentheses below regression coefficients, with statistical significance indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The Age^2 variable was divided by 1,000 to rescale its coefficient

Table 5 Heterogeneous treatment effects: Oppose Local Development

	Enviro	Non Enviro	Visited Ocean	Never visited Ocean
Sorry	0.30	0.32	0.40	0.34
	(0.374)	(0.352)	(0.559)	(0.293)
Blame	0.16	−0.17	0.35	−0.041
	(0.377)	(0.392)	(0.585)	(0.322)
Good	−0.70*	−0.47	0.86	−0.88***
	(0.411)	(0.362)	(0.680)	(0.312)
Good, Sorry	−0.56	−0.91**	0.27	−0.97***
	(0.440)	(0.411)	(0.674)	(0.363)
Good, Blame	−0.40	−0.50	−0.34	−0.42
	(0.373)	(0.347)	(0.633)	(0.295)
Bad	0.35	1.02***	0.32	0.92***
	(0.425)	(0.381)	(0.697)	(0.315)
Bad, Sorry	0.51	0.69*	0.066	0.79***
	(0.395)	(0.411)	(0.634)	(0.302)
Bad, Blame	0.87**	0.41	1.16	0.61**
	(0.423)	(0.373)	(0.882)	(0.304)
Visit ocean	0.25	0.20		
	(0.232)	(0.318)		
Visit park	0.35	−0.45**	0.18	−0.19
	(0.261)	(0.199)	(0.412)	(0.170)
Enviro	−0.57*	−0.42**	−0.73***	−0.65***
	(0.306)	(0.184)	(0.211)	(0.110)
Age	0.11**	0.061	0.17*	0.069**
	(0.0460)	(0.0395)	(0.0988)	(0.0315)
Age^2	−0.0011**	−0.00071*	−0.0018*	−0.00075**
	(0.000476)	(0.000407)	(0.00107)	(0.000326)
Kids	0.25	0.25	0.17	0.23
	(0.220)	(0.225)	(0.377)	(0.173)
Income	−0.11	−0.049	−0.078	−0.073
	(0.0753)	(0.0795)	(0.120)	(0.0582)
N	384	357	155	586
Pseudo R^2	0.036	0.035	0.049	0.055

Note. Respondents stated the likelihood that they would engage in each action on a seven-point Likert scale. Regressions are estimated by ordered logit, but results are robust to ordered probit and OLS estimation. Treatment dummies are relative to the baseline (No Apology, No Reputation) treatment. Robust standard errors are given in parentheses below regression coefficients, with statistical significance indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The Age^2 variable was divided by 1,000 to rescale its coefficient

Table 6 Heterogeneous treatment effects: Sign a Petition

	Enviro	Non Enviro	Visited Ocean	Never visited Ocean
Sorry	0.18	0.49	−0.20	0.60*
	(0.425)	(0.387)	(0.640)	(0.321)
Blame	−0.19	0.080	0.14	−0.041
	(0.409)	(0.376)	(0.594)	(0.319)
Good	−0.58	−0.17	0.71	−0.52
	(0.460)	(0.374)	(0.673)	(0.333)
Good, Sorry	−0.82*	−0.80**	−0.30	−0.84**
	(0.467)	(0.372)	(0.671)	(0.347)
Good, Blame	−0.65	−0.32	−0.65	−0.36
	(0.401)	(0.367)	(0.673)	(0.303)
Bad	0.46	1.55***	0.94	1.20***
	(0.448)	(0.378)	(0.735)	(0.323)
Bad, Sorry	0.49	0.52	−0.20	0.79**
	(0.425)	(0.406)	(0.625)	(0.315)
Bad, Blame	0.84**	0.60	1.43	0.74**
	(0.425)	(0.384)	(0.887)	(0.308)
Visit ocean	0.34	0.15		
	(0.224)	(0.335)		
Visit park	0.20	−0.50***	−0.11	−0.28
	(0.285)	(0.189)	(0.490)	(0.172)
Enviro	−0.61**	−0.39**	−0.80***	−0.67***
	(0.309)	(0.166)	(0.243)	(0.109)
Age	0.086*	0.014	0.20**	0.024
	(0.0468)	(0.0434)	(0.0900)	(0.0342)
Age^2	−0.00087*	−0.00029	−0.0021**	−0.00032
	(0.000485)	(0.000448)	(0.000935)	(0.000354)
Kids	0.089	0.0064	0.011	0.078
	(0.217)	(0.230)	(0.396)	(0.170)
Income	−0.087	−0.039	−0.042	−0.042
	(0.0719)	(0.0741)	(0.108)	(0.0577)
N	384	357	155	586
Pseudo R^2	0.037	0.037	0.067	0.053

Note. Respondents stated the likelihood that they would engage in each action on a seven-point Likert scale. Regressions are estimated by ordered logit, but results are robust to ordered probit and OLS estimation. Treatment dummies are relative to the baseline (No Apology, No Reputation) treatment. Robust standard errors are given in parentheses below regression coefficients, with statistical significance indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The Age^2 variable was divided by 1,000 to rescale its coefficient

Table 7 Treatment effects on preferred fine size

	OLS		Tobit		Ordered logit	
Sorry	0.25	0.41	0.27	0.48	0.081	0.15
	(0.864)	(0.842)	(0.920)	(0.890)	(0.275)	(0.279)
Blame	−0.55	−0.29	−0.58	−0.25	−0.23	−0.11
	(0.868)	(0.857)	(0.914)	(0.900)	(0.291)	(0.300)
Good	−0.30	−0.22	−0.32	−0.19	−0.099	−0.051
	(0.890)	(0.883)	(0.940)	(0.926)	(0.292)	(0.302)
Good, Sorry	−0.70	−0.40	−0.73	−0.41	−0.25	−0.14
	(0.916)	(0.930)	(0.962)	(0.967)	(0.307)	(0.322)
Good, Blame	−1.63*	−1.36	−1.70*	−1.40	−0.50*	−0.44
	(0.845)	(0.864)	(0.880)	(0.890)	(0.272)	(0.287)
Bad	1.67*	1.87**	1.81*	2.04**	0.50*	0.65**
	(0.878)	(0.890)	(0.959)	(0.964)	(0.281)	(0.297)
Bad, Sorry	1.15	1.24	1.20	1.35	0.37	0.47*
	(0.825)	(0.819)	(0.883)	(0.871)	(0.260)	(0.272)
Bad, Blame	1.06	1.49*	1.13	1.62*	0.33	0.49*
	(0.876)	(0.862)	(0.943)	(0.920)	(0.276)	(0.284)
Visit ocean		0.81*		0.87		0.31*
		(0.493)		(0.532)		(0.160)
Visit park		−0.76		−0.79		−0.30*
		(0.495)		(0.519)		(0.168)
Enviro		1.24***		1.34***		0.45***
		(0.263)		(0.281)		(0.0993)
Age		0.020		0.021		0.0050
		(0.0903)		(0.0963)		(0.0315)
Age^2		−0.78		−0.82		−0.25
		(0.920)		(0.974)		(0.324)
Kids		−0.49		−0.55		−0.18
		(0.477)		(0.510)		(0.165)
Income		0.0095		0.0084		−0.0056
		(0.135)		(0.144)		(0.0450)
Intercept	8.85***	13.4***	8.93***	13.8***		
	(0.626)	(2.336)	(0.667)	(2.506)		
N	741	741	741	741	741	741
R^2 / Pseudo R^2	0.030	0.088	0.0055	0.017	0.0086	0.027
Tobit σ			5.74***	5.56***		
			(0.157)	(0.153)		

Note. Treatment dummies are relative to the baseline (No Apology, No Reputation) treatment. Robust standard errors are given in parentheses below regression coefficients, with statistical significance indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The largest option for preferred fine size was "More than $15 million." OLS regressions top code this response at $17.5 million, while tobit regressions censor values at $15 million and above, and the ordered logit treats each potential response as a discrete choice with an ordinal ranking. The Age^2 variable was divided by 1,000 to rescale its coefficient

the relative ranking of the Full Apology and Blame treatments depends on whether control variables are included in the model. It is important to keep in mind that this fine is purely a punishment; in this section subjects were asked to consider the fine the firm should have to pay *in addition* to clean-up costs and stakeholder compensation.

In the treatments with no reputational information, the "Sorry" and "Blame" coefficients are imprecisely estimated but have opposite signs. This suggests that any kind of apology by itself has no effect on punishment, but that again, a "Shift the Blame" strategy is at least as effective as fully apologizing. Either type of apology is most effective when a firm has a good reputation, but is even slightly effective when the firm has a bad reputation, although neither of these effects is statistically different than its "reputation only" comparison. This may be because people expect apologies from bad actors.[13] Most surprising is that even with objectively bad information on the firm's reputation, the firm's signals (apology or blame) seem to weakly reduce the preferred fine so that it is no longer statistically different than the baseline treatment.

Although the largest impacts come from objective reputational information, interestingly an objectively good reputation does not influence the fine much unless it's combined with a signal — specifically a signal that shifts the blame rather than offers a full apology. Conditional on a good reputation, shifting the blame causes a reduction in the preferred fine of between $1.4 and $1.7 million, although this is only statistically significant at the 10% level. On the other hand, bad reputational information has the biggest impact of all information types, leading to preferred fines that are between $1 and $2 million dollars more than in the baseline treatment.

In summary, signals can influence preferred fine sizes only slightly in the presence of bad objective information, but good objective information must be combined with a signal from the firm in order to influence households, and blame signals are at least as effective as apologies unless the firm already has a bad reputation. These results suggest that blaming another party may sow doubt in the subject's mind about how severely they should punish the offender, which has implications for how juries may award damages in court.

4.3 Willingness to accept

Estimates from the Willingness to Accept model Eqs. 3–5 are given in Table 8. The first two columns report the effect of treatment on the probability of accepting a given bid, with and without control variables and with the baseline treatment (No Apology, No Information) as the omitted dummy variable. The bid coefficient is positive and statistically significant as expected. The coefficient on the "Sorry" (Full Apology, No Information), and "Bad, Blame" treatments are statistically significant at the 10 percent level and of comparable magnitude, suggesting that these treatments make respondents on average less likely to accept a given compensation bid. None of the

[13] In criminal trial settings, for example, many jurisdictions require statements of remorse from violent criminals upon conviction and before sentencing.

Table 8 Estimating WTA with double-bounded dichotomous choice bids

	MLE Dependent Variable: WTA Bid Yes/No		WTA Estimates Median	Mean
Baseline			38.6	169
			(26.8, 54.5)	(101, 452)
Sorry	−0.51*	−0.54*	59.7	262
	(0.303)	(0.302)	(42.8, 82.6)	(151, 766)
Blame	−0.34	−0.35	51.3	225
	(0.280)	(0.284)	(38.0, 68.5)	(136, 621)
Good	−0.25	−0.32	50.1	220
	(0.291)	(0.297)	(35.7, 69.3)	(130, 602)
Good, Sorry	−0.21	−0.27	48.0	211
	(0.298)	(0.310)	(33.8, 68.6)	(125, 593)
Good, Blame	−0.25	−0.28	48.3	212
	(0.293)	(0.297)	(34.8, 66.6)	(125, 614)
Bad	−0.15	−0.13	42.7	187
	(0.293)	(0.301)	(30.6, 59.7)	(112, 509)
Bad, Sorry	−0.36	−0.36	51.5	226
	(0.290)	(0.297)	(37.7, 71.4)	(135, 639)
Bad, Blame	−0.51*	−0.54*	59.6	262
	(0.291)	(0.295)	(43.5, 81.9)	(154, 734)
Bid	1.20***	1.24***		
	(0.0775)	(0.0797)		
Intercept	−4.38***	−4.26***		
	(0.381)	(0.879)		
Control variables	No	Yes	Yes	Yes
N	741	741		
LL	−909	−890		

Note. Treatment dummies are relative to the baseline (No Apology, No Information) treatment. MLE represents maximum likelihood estimation. Robust standard errors are given in parentheses below regression coefficients, with statistical significance indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The 95% confidence intervals given in parentheses below the WTA estimates were estimated using the method of Krinsky and Robb (1986), based on the model in column 2 (including control variables). Control variable coefficients are suppressed for brevity but are available in the corresponding online appendix. Control variables were: ocean and national park visitation dummies, environmental sentiment, age, age squared, income, and children dummy

other treatment effects have a statistically significantly different probability than the baseline treatment of accepting a given bid. However, the Bad Reputation treatment (with no apology message) is numerically closest to zero, suggesting that subjects are least likely to seek compensation from a notoriously bad actor. Subjects may be weighing strategic aspects of the bid setting along with their desire to punish the

firm as well as their own WTA for lost passive use value, which may explain the nonmonotonic pattern in WTA estimates across apology types and reputation quality shown in Table 8 and Fig. 1.

We used the coefficients and covariance matrix from the model in column 2 in order to estimate the median and mean WTA for each treatment group and to simulate the 95% confidence intervals with the Krinsky and Robb (1986) method. The last two columns of Table 8 give the mean and median WTA estimates and Fig. 1 gives box plots of simulated distributions. While these two measures of expected WTA are not statistically different across treatments, the estimates qualitatively mirror the regression results; there seems to be no strong case that an apology reduces the compensation demanded by stakeholders. The "Sorry" and "Bad, Blame" treatments induce not only the largest compensation demanded, but also the greatest variance in the estimated mean and median.

5 Discussion

Our results illustrate the importance of both firm reputation and a genuine apology in the aftermath of a major man-made environmental disaster. We find that the reputational effects are especially strong. The public says they are significantly more likely to boycott an offending firm if it has a history of culpability and a reputation of being

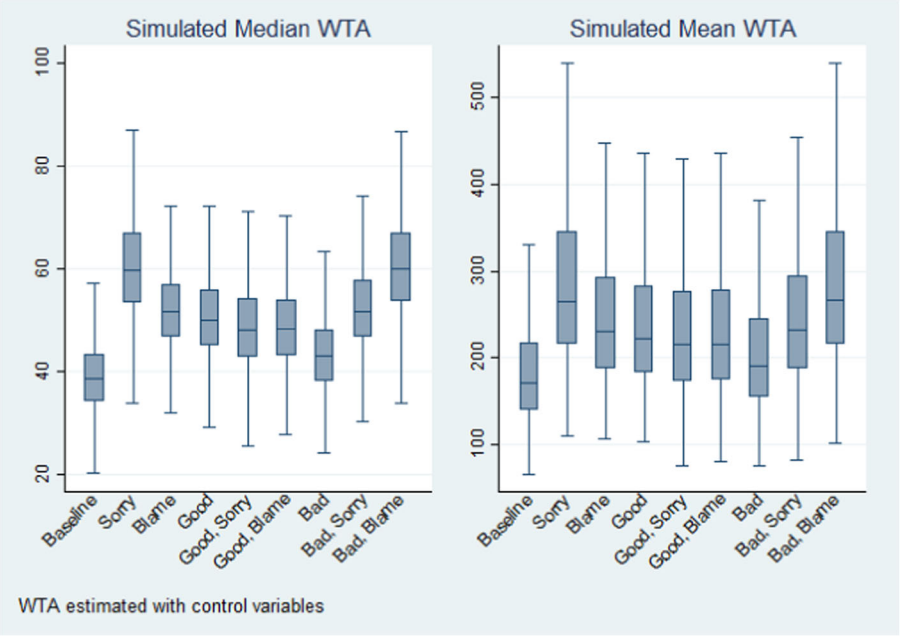

Fig. 1 Distribution of WTA

a bad actor. A firm with a bad reputation, however, can issue a genuine apology to help soften the public's desire to punish the firm. For firms with a good reputation, we find that a genuine apology combined with shifting the blame to another agent may be an even more effective strategy to reduce punishment. This is in contrast to the classic theory of apologies as costly signals, in which an apology without cost has no value at all (see e.g., Ho 2012). Here we find a case in which an apology without costs has value—when a firm has a good reputation, they can lower expected fines by showing remorse and then shifting blame to another actor. This suggests an effective apology strategy is twofold—a bad reputational firm should offer a sincere apology and accept fault, while a good reputational firm should express remorse and shift blame towards others.

Consider first a bad reputation example—the Exxon Valdez oil spill. On March 24, 1989 an Exxon Valdez oil tanker ran aground in Prince William Sound, Alaska, releasing about 11 million gallons of crude oil into the bay. Exxon's response is a case study in how not to apologize. First, Exxon's Chief Executive Officer sent lower level executives to the site, and he waited six days before making his first public statement. Three additional days passed before Exxon apologized officially by running an advertisement in 160 newspapers—but the firm did not accept responsibility and readers found the tone of the apology to be insincere. Their response was made even more tone deaf when one Exxon executive suggested that if the company was forced to pay for the clean-up, they would raise gas prices to finance the effort. The public viewed Exxon's apology as insincere—it was delayed, offered in print (executives were not on TV to visually express sincerity), and was not viewed as costly or fair (see O'Hara O'Connor 2011). This botched response was widely criticized, and arguably contributed to the subsequent decline in Exxon's stock price (O'Hara O'Connor 2011).

Now consider a "good" reputation example—the BP Gulf of Mexico spill. Roughly 4,000 miles away from Prince William Sound, and two decades after the Exxon Valdez spill, an explosion occurred on an oil drilling rig owned by Transocean and leased by British Petroleum (BP). The well leak released nearly 5 million barrels of crude oil into the Gulf (Silliman et al. 2012). But BP had a relatively good reputation prior to the April 2010 spill due to several strategic "green" decisions. Before the spill, BP invested about $200 million on a green advertising campaign termed "Beyond Petroleum". The company re-branded its gas stations with the Helios symbol and renamed its BP acronym to "Beyond Petroleum". The new BP name was "designed to reflect the company's newly stated dedication to environmental stewardship" (Barrage, Chyn, and Hastings 2014). This pre-spill advertising campaign could have affected the public's beliefs about whether the subsequent oil spill was due to negligence or bad luck befalling a green firm (Minor and Morgan 2011). Consistent with our findings, evidence suggests the BP campaign effectively softened the consumer response to the oil spill (Barrage, Chyn, and Hastings 2014). Given BP's green branding efforts, our results also suggest that in response to the spill, BP should have apologized but also should have stopped short of accepting responsibility. In fact, that is what happened: following the spill, BP executives quickly expressed remorse on camera, and did so with emotion. CEO Tony Hayward personally visited the affected zones, and the company offered $20 billion to help compensate affected parties

(Barrage, Chyn, and Hastings 2014). But the company also blamed Transocean (the owner of the drilling rig) for the explosion, resulting fatalities, and environmental damage.

6 Conclusion

Following major environmental disasters, such as those that occurred in Valdez, Alaska in 1989 and the Gulf of Mexico in 2010, transgressing firms sometimes offer apologies that vary in terms of completeness. Firms also vary according to their reputation; some have a history of environmental and/or safety infractions whereas others have a more responsible record.

We study the importance of both apologies and firm reputation following a firm-caused environmental disaster. We presented subjects with a fictional environmental disaster scenario (an oil spill off the coast of California) and varied both firm reputation and type of apology that was offered by the transgressing firm. We measure treatment effects on 1) individual responses, e.g., the reported likelihood that a subject will oppose the firm in some way in the future, 2) the preferred size of fine that is charged to the firm and 3) the average willingness to accept compensation for the resulting environmental damage.

Relative to the baseline (No Apology, No Reputation) treatment, a good reputation decreases the likelihood of individual action (e.g., boycotting the offending firm) whereas a bad reputation has the opposite effect. A full apology amplifies the effect of a good reputation and reduces the effect of a bad reputation, whereas a partial apology (i.e., shifting the blame) is relatively less effective at reducing the likelihood of individual action.

When a firm has a bad reputation, people tend to ask for a larger fine, and this effect can be mitigated some by offering any type of apology (accepting responsibility or shifting the blame). When a firm has a good reputation, shifting the blame is actually most effective at reducing the preferred fine size. We also find that offering a full apology raises subjects' willingness to accept for the corresponding damage. These results suggest that offering a complete apology may not be optimal as doing so effectively admits guilt—and this results in heftier fines and additional compensation demanded.

Taken all together, one speculative interpretation of our results is that people view things like compensation and associated fines as due consequences of infractions, but view individual actions (such as boycotts) as remedies for callous behavior. In other words, people want guilty actors to be held accountable, but don't take vengeful actions unless they feel the actor is not accepting responsibility for their actions or is otherwise not being held accountable. This suggests that a firm should optimally offer as sincere of an apology as possible, but stop short of accepting blame. While this interpretation might help explain why so many corporate apologies tend to be partial ones, more research ultimately needs to be carried out to better understand when a partial apology is strategically preferred to a full one. Along these lines, incorporating apologies into public goods games with willing punishers (e.g., Fehr and Gächter 2000, 2002), in which free riders are given the opportunity

to apologize to potential punishers, may help explain why and when apologies matter.

Appendix A: Perceived versus actual effect of apologies and reputation information

We asked subjects who received an apology treatment whether they thought the apology affected their valuation decisions, and if so in what way — a smaller fine or WTA, a larger one, or neither. We also asked a similar question for the information treatments. In this appendix we compare these responses to the actual treatment effects, separately estimating the effects for the subgroup who felt the apology or reputation information influenced them.

Table 9 compares these results for the apology treatments, where treatment dummies are defined relative to the "Sorry" treatment (Full Apology only, No Reputation). The first column shows that good or bad reputation information makes subjects more or less likely to believe the apology affected their decisions, but that blame shifting does not have this effect. The second and third columns show that, among subjects who believe the apology affected their decisions, those receiving good reputation information also *believe* they are less likely to demand a larger fine. Consistent with the "size of fine" results in Table 7 discussed above, this effect is larger when the "Shift the Blame" message is paired with a good reputation. However, the results are much stronger here, in column 3 of Table 9, for the *perceived* effect on a subject's preferred fine among subjects who believe the apologies matter to them. Compare this perceived effect to column 6 of Table 9 which separates the treatment effects on preferred fine size by whether or not the subject believes the apology mattered to them. In this column we reestimate (2) with an interaction between treatment dummies and a dummy that is equal to one if the subject stated the apology affected their decisions. In this column, the baseline group consists of subjects in the "Sorry" (Full Apology) treatment with no reputation information, who do not believe the apology mattered to them. Relative to that group, subjects in the same treatment group who believe the apology affected them demanded a $3.48 million *smaller* fine on average. However any additional reputation information works in the opposite direction for this group to partially undo the extent of their forgiveness. So while the apology-affected group did not demand larger fines on average or in any specific treatment, consistent with column 3, they seem to extend additional judgments when more information is provided. Notably, the "Good, Blame" treatment which we noted earlier when discussing Table 7 was most effective at reducing the demanded fine is most effective among the group who believes apologies did not matter to them; this group demanded a $2.26 million smaller fine in this treatment.

Similarly, the fourth and fifth columns of Table 9 show that subjects who believe the apology influenced them also say they are more likely to demand less compensation if the apology is combined with a good reputation, and less likely to demand more compensation if a firm with a bad reputation apologizes. Yet we saw in Table 8 that full apologies combined with reputation information had almost no effect on compensation demanded. We again compare this perceived effect in columns 4 and

Table 9 Comparing perceived effect of apology to actual treatment effects

	Debriefing questions					Valuation questions	
	Logit	Multinomial logit		Multinomial logit		Tobit	MLE
	Did apology matter	Apol. affect preferred fine		Apol. affect WTA		Preferred fine size	WTA Bid
		Lower	Higher	Lower	Higher		
Blame	0.054	1.03	0.38	1.4	0.21	−1.43	0.49
	(0.331)	(0.803)	(0.739)	(0.890)	(0.763)	(1.016)	(0.326)
Good, Sorry	0.60*	−0.11	−2.21***	1.96**	−1.35	−1.17	0.84**
	(0.337)	(0.756)	(0.794)	(0.838)	(1.169)	(1.320)	(0.402)
Good, Blame	0.51	0.76	−3.31**	1.11	−1.95	−2.26**	0.32
	(0.339)	(0.675)	(1.305)	(0.824)	(1.246)	(1.071)	(0.364)
Bad, Sorry	−1.02***	−1.02	−1.34	−0.68	−2.58*	0.18	0.43
	(0.375)	(1.265)	(1.193)	(1.537)	(1.375)	(0.937)	(0.332)
Bad, Blame	−0.54	0.32	−0.56	0.52	0.97	0.059	0.18
	(0.379)	(0.918)	(0.936)	(1.197)	(0.990)	(0.992)	(0.329)
Apol. Matter						−3.48***	1.14**
						(1.150)	(0.518)
Apol. Matter X						2.12	−1.13*
Blame						(1.678)	(0.663)
Apol. Matter X						1.73	−1.69**
Good, Sorry						(1.773)	(0.685)
Apol. Matter X						2.00	−0.56
Good, Blame						(1.651)	(0.658)

Table 9 (continued)

Apol. Matter X						1.10	−0.81
Bad, Sorry						(1.733)	(0.718)
Apol. Matter X						3.51*	−0.49
Bad, Blame						(1.966)	(0.789)
Intercept	0.77	0.9	1.6	−1.9	−0.73	15.4***	−4.63***
	(1.173)	(2.532)	(2.824)	(2.821)	(3.063)	(2.814)	(1.118)
N	502	146		132		502	502
Pseudo R^2	0.089	0.27		0.28		0.026	
Tobit σ						5.33***	
						(0.179)	

Note. This table compares responses to debriefing questions about whether and how an apology mattered, to the responses given in the valuation tasks. Treatment dummies are relative to the Full Apology only, No Reputation information treatment. In the last two columns, which investigate the valuation tasks, treatment dummies are interacted with a dummy for whether the respondent believed the apology affected their valuations. All specifications included control variables. Coefficients are suppressed for brevity but are available in the corresponding online appendix. Control variables were: ocean and national park visitation dummies, environmental sentiment, age, age squared, income, and children dummy. Regressions with debriefing questions also control for preferred fine size and WTA. Robust standard errors are given in parentheses below regression coefficients, with statistical significance indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The Age^2 variable was divided by 1,000 to rescale its coefficient

Table 10 Comparing perceived effect of reputation information to actual treatment effects

	Debriefing questions					Valuation questions	
	Logit	Multinomial logit		Multinomial logit		Tobit	MLE
	Did info	Info affect preferred fine		Info affect WTA		Preferred	WTA
	Matter	Lower	Higher	Lower	Higher	Fine size	Bid
Good, Sorry	0.59	−0.047	−1.44**	0.31	−0.78	0.48	0.45
	(0.380)	(0.540)	(0.723)	(0.540)	(0.714)	(1.872)	(0.485)
Good, Blame	0.59	−0.25	−0.83	0.0027	−0.49	−1.19	0.70
	(0.368)	(0.525)	(0.637)	(0.505)	(0.665)	(1.583)	(0.516)
Bad	0.49	−16.2***	0.98*	−16.2***	0.62	−0.96	1.31**
	(0.366)	(0.543)	(0.594)	(0.495)	(0.584)	(1.641)	(0.612)
Bad, Sorry	0.38	−3.38***	0.95	−2.28***	0.68	−1.22	1.97***
	(0.330)	(1.185)	(0.581)	(0.871)	(0.579)	(1.538)	(0.509)
Bad, Blame	0.53	−3.13***	1.30**	−3.15***	1.04*	−0.38	0.89*
	(0.358)	(1.149)	(0.596)	(1.138)	(0.572)	(1.750)	(0.477)
Info Matter						−2.25*	1.14***
						(1.313)	(0.418)
Info Matter X						−0.59	−0.67
Good, Sorry						(2.176)	(0.621)
Info Matter X						0.37	−1.03
Good, Blame						(1.875)	(0.627)
Info Matter X						4.73**	−1.63**
Bad						(1.991)	(0.706)
Info Matter X						4.20**	−2.85***
Bad, Sorry						(1.835)	(0.617)
Info Matter X						3.17	−1.67***
Bad, Blame						(2.046)	(0.595)
Intercept	0.52	−1.82	−2.67	2.69	0.87	16.4***	−5.54***
	(1.153)	(2.095)	(1.860)	(2.351)	(2.163)	(3.021)	(1.180)
N	478	329		309		478	478
Pseudo R^2	0.044	0.37		0.32		0.022	
Tobit σ						5.39***	
						(0.189)	

Note. This table compares responses to debriefing questions about whether and how reputation information mattered, to the responses given in the valuation tasks. Treatment dummies are relative to the Good Reputation, No Apology treatment. In the last two columns, which investigate the valuation tasks, treatment dummies are interacted with a dummy for whether the respondent believed the information affected their valuations. All specifications included control variables. Coefficients are suppressed for brevity but are available in the corresponding online appendix. Control variables were: ocean and national park visitation dummies, environmental sentiment, age, age squared, income, and children dummy. Regressions with debriefing questions also control for preferred fine size and WTA. Robust standard errors are given in parentheses below regression coefficients, with statistical significance indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5 of Table 9 to column 7. As with the fine size, in this column we reestimate (3) with an interaction between treatment dummies and a dummy for whether or not the subject believed the apology mattered to them. Subjects in the "Sorry" treatment with no reputational information were significantly more likely to accept a given compensation bid if they believe apologies matter - which suggests there is considerable heterogeneity in the response given that the average effect of the "Sorry" treatment (Table 8) was to *increase* WTA. However, as with the fine size, this effect is at least partially offset by any additional reputation information or alternative apology message, as the interactions of the treatment dummies with the "Apology Matter" dummy show.

Table 10 shows a similar cognitive dissonance between the perceived and actual effect of reputation information. The first column shows that treatments have no effect on whether subjects think the information influenced their decisions. The second and third columns show that subjects who believe they responded to reputation information also say that they are less likely to advocate for a lower fine if the firm has a bad reputation unless the firm offers some form of apology. This *is* consistent with the average treatment effects on the preferred fine described in Table 7 and the ranking of interaction coefficients in the second to last column of Table 10. However, the fourth and fifth columns of Table 10 show that subjects think they behave in the same way regarding compensation, which is clearly not consistent with the last column of interaction coefficients on the likelihood of accepting a given bid, or results from Table 8.

References

Abeler, J., Calaki, J., Andree, K., Basek, C. (2010). The power of apology. *Economics Letters*, *107*(2), 233–235.

Barrage, L., Chyn, E., Hastings, J. (2014). Advertising, reputation, and environmental stewardship: Evidence from the BP oil spill. *Ann Arbor*, *1001*, 48109–1220.

Behar, R. (1990). Interview with Lawrence Rawl: Exxon Strikes Back. Time Magazine. http://www.time.com/time/magazine/article/0,9171,969673,00.html. Accessed 1 March 2017.

Benoit, W.L. (1997). Image repair discourse and crisis communication. *Public Relations Review*, *23*(2), 177–186.

Boothman, R.C., Blackwell, A.C., Campbell, D.A. Jr, Commiskey, E., Anderson, S. (2009). A better approach to medical malpractice claims? The University of Michigan experience. *Journal Health of Life Sciences Law*, *2*(2), 125–159.

Bottom, W.P., Gibson, K., Daniels, S.E., Murnighan, J.K. (2002). When talk is not cheap: Substantive penance and expressions of intent in rebuilding cooperation. *Organization Science*, *13*(5), 497–513.

Cherry, M.A., & Sneirson, J.F. (2011). Beyond profit: Rethinking corporate social responsibility and greenwashing after the BP oil disaster . *Tulane Law Review*, *85*(4), 983–1038.

Cohen, J.R. (1999). Apology and organizations: Exploring an example from medical practice. Fordham. *Urban Law Journal*, *27*, 1447.

De Cremer, D. (2010). To pay or to apologize? On the psychology of dealing with unfair offers in a dictator game. *Journal of Economic Psychology*, *31*(6), 843–848.

De Cremer, D., van Dijk, E., Pillutla, M.M. (2010). Explaining unfair offers in ultimatum games and their effects on trust: An experimental approach. *Business Ethics Quarterly*, *20*(01), 107–126.

De Cremer, D., Pillutla, M.M., Folmer, C.R. (2011). How important is an apology to you? Forecasting errors in evaluating the value of apologies. *Psychological Science*, *22*(1), 45–48.

Delmas, M.A., & Burbano, V.C. (2011). The drivers of greenwashing. *California Management Review*, *54*(1), 64–87.

Deloitte. (2014). Global survey on reputation risk. http://www2.deloitte.com/content/dam/Deloitte/pl/Documents/Reports/pl_Reputation_Risk_survey_EN.pdf. Accessed 1 March 2017.

Desmet, P.T., De Cremer, D., van Dijk, E. (2011). In money we trust? The use of financial compensations to repair trust in the aftermath of distributive harm. *Organizational Behavior and Human Decision Processes*, *114*(2), 75–86.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137.

Fischbacher, U., & Utikal, V. (2013). On the acceptance of apologies. *Games and Economic Behavior*, *82*, 592–608.

Gold, G.J., & Weiner, B. (2000). Remorse, confession, group identity, and expectancies about repeating a transgression. *Basic and Applied Social Psychology*, *22*(4), 291–300.

Haab, T.C., & McConnell, K.E. (2002). *Valuing environmental natural resources: The econometrics of non-market valuation*. Northampton: Edward Elgar Publishing.

Hearit, K.M. (2006). Crisis management by apology: Corporate response to allegations of wrongdoing. Routledge.

Ho, B. (2012). Apologies as signals: With evidence from a trust game. *Management Science*, *58*(1), 141–158.

Ho, B., & Liu, E. (2011). Does sorry work? The impact of apology laws on medical malpractice. *Journal of Risk and Uncertainty*, *43*(2), 141–167.

Jeanty. P.W. (2007). wtpcikr: Constructing Krinsky and Robb confidence interval for mean and median willingness to pay (wtp) using stata. North American Stata Users' Group Meetings, August 2007.

Kim, P.H., Ferrin, D.L., Cooper, C.D., Dirks, K.T. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence-versus integrity-based trust violations. *Journal of Applied Psychology*, *89*(1), 104.

Knight, R.F., & Pretty, D.J. (1996). The impact of catastrophes on shareholder value. Templeton College.

Kotchen, M., & Moon, J.J. (2012). Corporate social responsibility for irresponsibility. *The BE Journal of Economic Analysis and Policy*, *12*(1), 1–23.

Korobkin, R., & Guthrie, C. (1994). Psychological barriers to litigation settlement: An experimental approach. *Michigan Law Review*, *93*(1), 107–192.

Krinsky, I., & Robb, A.L. (1986). On approximating the statistical properties of elasticities. *The Review of Economic and Statistics*, *68*(4), 715–719.

Loomis, J., & Ekstrand, E. (1998). Alternative approach for incorporating uncertainty when estimating willingness to pay: The case of the Mexican spotted owl. *Ecological Economics*, *27*, 29–41.

Martinez-Vaquero, L.A., Han, T.A., Pereira, L.M., Lenaerts, T. (2015). Apology and forgiveness evolve to resolve failures in cooperative agreements. *Scientific Reports*, *5*(10639), 1–12.

McMichael, B.J., Van Horn, R.L., Viscusi, W.K. (2016). Sorry is never enough: The effect of state apology laws on medical malpractice liability risk. Available at SSRN: https://ssrn.com/abstract=2883693 or https://doi.org/10.2139/ssrn.2883693.

Minor, D., & Morgan, J. (2011). CSR as reputation insurance: Primum non nocere. *California Management Review*, *53*(3), 40–59.

Mungan, M.C. (2012). Don't say you're sorry unless you mean it: Pricing apologies to achieve credibility. *International Review of Law and Economics*, *32*(1), 178–187.

O'Hara O'Connor, E.A. (2004). Apology and thick trust: What spouse abusers and negligent doctors might have in common. *Chicago-Kent Law Review*, *79*, 1055–1089.

O'Hara O'Connor, E.A. (2011). Organizational apologies: BP as a case study. *Vanderbilt Law Review*, *64*(6), 1959–1991.

O'Hara O'Connor, E.A., & Yarn, D. (2002). On apology and consilience. *Washington Law Review*, *77*, 1121–1192.

O'Hara O'Connor, E.A., & Robbins, M.M. (2009). Using criminal punishment to serve both victim and social needs. *Law and Contemporary Problems*, *72*(2), 199–217.

Ohtsubo, Y., & Watanabe, E. (2009). Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior*, *30*(2), 114–123.

Pace, K.M., Fediuk, T.A., Botero, I.C. (2010). The acceptance of responsibility and expressions of regret in organizational apologies after a transgression. *Corporate Communications: An International Journal*, *15*(4), 410–427.

Park, T., Loomis, J.B., Creel, M. (1991). Confidence interval for evaluating benefit estimates from dichotomous choice contingent valuation studies. *Land Economics*, *67*, 64–73.

Pauly, J.J., & Hutchison, L.L. (2005). Moral fables of public relations practice: The Tylenol and Exxon Valdez cases. *Journal of Mass Media Ethics*, *20*(4), 231–249.

Poe, G.L., Giraud, K.L., Loomis, J.B. (2005). Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics*, *87*, 353–365.

Robbennolt, J.K. (2003). Apologies and legal settlement: An empirical examination. *Michigan Law Review*, *102*(3), 460–516.

Schniter, E., Sheremeta, R.M., Sznycer, D. (2013). Building and rebuilding trust with promises and apologies. *Journal of Economic Behavior and Organization*, *94*, 242–256.

Silliman, B.R., van de Koppel, J., McCoy, M.W., Diller, J., Kasozi, G.N., Earl, K., Adams, P.N., Zimmerman, A.R. (2012). Degradation and resilience in Louisiana salt marshes after the BP-Deepwater Horizon oil spill. *Proceedings of the National Academy of Sciences*, *109*(28), 11234–11239.

Small, W.J. (1991). Exxon Valdez: How to spend billions and still get a black eye. *Public Relations Review*, *17*(1), 9–25.

Spence, M. (1973). Job market signaling. *The Quarterly Journal Of Economics*, *87*(3), 355–374.

Tyler, L. (1997). Liability means never being able to say you're sorry: Corporate guilt, legal constraints, and defensiveness in corporate communication. *Management Communication Quarterly*, *11*(1), 51–73.

Williams, D.E., & Treadaway, G. (1992). Exxon and the Valdez accident: A failure in crisis communication. *Communication Studies*, *43*(1), 56–64.